



HOCHSCHULE  
**harz**

Hochschule für angewandte  
Wissenschaften (FH)

Wernigerode  
Halberstadt



Frieder Stolzenburg (Hrsg.)

2. Wernigeröder  
Automatisierungs- und Informatiktage  
Hochschule Harz, Wernigerode

Tagungsband

**HARZER HOCHSCHULTEXTE**

**Nr.8**

## **Impressum**

2. Wernigeröder Automatisierungs- und Informatiktage

an der Hochschule Harz (FH) in Wernigerode, 12./13. Oktober 2006

Hrsg.: Prof. Dr. Frieder Stolzenburg

© 2006

Hochschule Harz (FH)

Fachbereich Automatisierung und Informatik

Friedrichstraße 57 – 59

38855 Wernigerode

Tel.: 03943/659-300

E-Mail: [fstolzenburg@hs-harz.de](mailto:fstolzenburg@hs-harz.de)

Redaktion: Programmkomitee der 2. Wernigeröder Automatisierungs- und  
Informatiktage

Harzer Hochschultexte Nr. 8

ISSN 1619-7232

Für den Inhalt und die Richtigkeit der einzelnen hier veröffentlichten Beiträge übernimmt der Herausgeber keine Verantwortung. Jeder Verfasser ist als Autor selbst für seinen Beitrag verantwortlich.

# Inhaltsverzeichnis

<b>Inhaltsverzeichnis</b> .....	<b>003</b>
<b>Vorwort</b> .....	<b>007</b>
<b>Organisation</b> .....	<b>009</b>
<b>I     <b>Eingeladener Beitrag I</b></b> .....	<b>011</b>
Global Parking Space Management: The Case for Database Web Services .....	013
<b>II    <b>IT-Sicherheit und E-Gouvernement</b></b> .....	<b>033</b>
Automatisierte Kopieerkennung am Beispiel Eintrittskarte ....	035
Enterprise Dokument Security .....	043
eGovernment für Hochschulen und Studierende – Herausforderungen und Möglichkeiten auch im Bologna-Prozess .....	051
Dezentrales Bewertungskonzept für Web Services .....	059
<b>III   <b>Informatik / Wissen für die Zukunft</b></b> .....	<b>063</b>
Management von Serviceprozessen im HCC Magdeburg auf Basis des ITIL-Frameworks .....	065
Zufallszahlen in der Lehre .....	073
Uniform random binary floating point number generation .....	077
Augmented Learning in der optischen Nachrichtentechnik – Konzeption einer interaktiven Lernsoftware zur Unter- stützung der selbstständigen Laborarbeit .....	085
<b>IV    <b>Eingeladener Beitrag II</b></b> .....	<b>091</b>
Wissen, Ziele, Aktionen – Eine unabdingbare Folge Auch in der Automation .....	093
<b>V     <b>Automatisierungstechnik I</b></b> .....	<b>103</b>
Proxykonzept zur Integration von intelligenten Feldgeräten in Steuerungssysteme .....	105
Vorschlag für ein Prozessinformationsdisplay zur Übersichtsdarstellung komplexer Prozesszustände .....	113
Intelligente Datenanalyse für die optische 3D-Inspektion .....	121
Lösungsansatz für eine ereignisgesteuerte Prozessdiagnose .....	129
<b>VI    <b>Automatisierungstechnik II</b></b> .....	<b>137</b>
Die umweltgerechte Gestaltung von Motoren und deren Bewertung .....	139
Selektiver Gassensor für Hochtemperaturanwendungen auf der Basis piezoelektrischer Resonatoren .....	147
Elektromechanische Eigenschaften von Langasit- resonatoren bei hohen Temperaturen .....	151

	Mikroelektromechanische Systemkomponenten für Hochtemperaturanwendungen .....	155
<b>VII</b>	<b>Mobile und Geoinformationssysteme .....</b>	<b>159</b>
	Autonomous Mobile Robots for Outdoor Tasks .....	161
	lejON – LeJOS Odometric Navigator .....	169
	Optimierung von mobilen Geosensornetzwerken unter Berücksichtigung sowohl der Phänomen- als auch Geosensornetzwerk-Charakteristika .....	177
	Praxistauglichkeitstests mit einem mobilen Geodatenerfas- sungssystem für einen tragbaren Xybernat-Computer .....	185





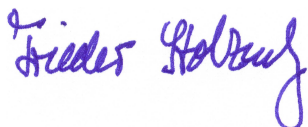
## Vorwort

Die 2. Wernigeröder Automatisierungs- und Informatiktage (kurz: WAIT'06) an der Hochschule Harz in Wernigerode liefern ein Forum zur Präsentation von Ergebnissen der angewandten Forschung in den Bereichen Automatisierung und Informatik. Beide durchdringen Wirtschaft und Verwaltung, aber auch den Alltag mehr und mehr. Kein Bereich ist heute ohne den anderen denkbar. Zielsetzung der Tagung ist es darum, die Forschungsaktivitäten und –ergebnisse gerade an der Schnittstelle zwischen Automatisierung und Informatik darzustellen, zu diskutieren und kritisch zu hinterfragen. So finden sich dazu im fast zweitägigem Tagungsprogramm, das noch durch eine Abendveranstaltung im Wernigeröder Rathaus abgerundet wird, eine ganze Bandbreite von Beiträgen – von Aspekten der Automatisierungstechnik über Mobile und Geoinformationssysteme bis hin zu IT-Sicherheit und E-Government.

Die Vorgängerveranstaltung dieser Tagung fand im Oktober 2002 ebenfalls an der Hochschule Harz statt und markierte gleichzeitig das 10-jährige Jubiläum des Fachbereichs Automatisierung und Informatik, der im Herbst 1992 gegründet wurde. Das Studienangebot des Fachbereiches Automatisierung und Informatik orientiert sich an der Anwendung der Informatik, der Telekommunikation und Automatisierungstechnik in allen Bereichen der Wirtschaft und Verwaltung. Er bietet die fünf Bachelorstudiengänge Intelligente Automatisierungssysteme, Informatik, Medieninformatik, Wirtschaftsinformatik und Wirtschaftsingenieurwesen sowie den Masterstudiengang Informatik/Mobile Systeme an. Durch das regelmäßig stattfindende Forum für Antriebssysteme und den Aufbau eines Kompetenzzentrums für Informations- und Kommunikationstechnologien profiliert sich der Fachbereich darüber hinaus in der angewandten Forschung auf den Gebieten der Automatisierung und Informatik.

Neben Beiträgen von der Hochschule Harz selbst ist es dank des öffentlichen Aufrufs nach Beiträgen zur WAIT'06 gelungen, Beiträge aus Hochschulen und Unternehmen aus ganz Deutschland zu gewinnen. Um ein hohes Niveau der Beiträge zu gewährleisten, wurden sämtliche Beiträge durch Mitglieder des 17-köpfigen Programmkomitees und weitere Wissenschaftler begutachtet. So konnten aus 25 Einreichungen 13 als Vollbeiträge und 7 als Kurzbeiträge mit Posterpräsentation in das Tagungsprogramm bzw. in den vorliegenden Tagungsband, der den Teilnehmern bereits zur Tagung selbst zur Verfügung steht, aufgenommen werden. Das Programm wird abgerundet durch zwei eingeladene Beiträge zu Datenbank-Web-Diensten und zu Zieldefinitionen für die Automation, die das Spannungsfeld zwischen Automatisierung und Informatik darstellen. Insofern darf ich auf rege Diskussionen und eine für alle Beteiligten erfolgreiche Tagung hoffen.

Wernigerode, im September 2006



Prof. Dr. Frieder Stolzenburg

Tagungsleiter der 2. Wernigeröder Automatisierungs- und Informatiktage  
Hochschule Harz (FH)





## Organisation

- Tagungsleitung:

Prof. Dr. Frieder Stolzenburg

- Programmkomitee:

Prof. Dr. Jana Dittmann, Universität Magdeburg  
Prof. Dr. Alexander Fay, HSU Hamburg  
Prof. Dr. Ulrich Fischer-Hirchert, Hochschule Harz (FH)  
Prof. Dr. Walter Gießler, Hochschule Harz (FH)  
Prof. Dr. Hartmut Hensel, Hochschule Harz (FH)  
Prof. Dr. Ulrich Jumar, ifak Magdeburg  
Prof. Dr. Klaus-Dietrich Kramer, Hochschule Harz (FH)  
Prof. Dr. Steven Liu, TU Kaiserslautern  
Dr. Bernd Markus Pfeiffer, Siemens Karlsruhe  
Prof. Dr. Hardy Pundt, Hochschule Harz (FH)  
Prof. Dr. Andreas Schmietendorf, FHW Berlin  
Prof. Dr. René Simon, Hochschule Harz (FH)  
Prof. Dr. Frieder Stolzenburg, Hochschule Harz (FH)  
Prof. Dr. Hermann Strack, Hochschule Harz (FH)  
Prof. Dr. Gerhard Weber, MMC Kiel  
Dr. Christian Zeidler, ABB Forschungszentrum, Ladenburg  
Prof. Dr. Bernhard Zimmermann, Hochschule Harz (FH)

- Weitere Gutachter

Dr. Martin Hollender  
Dr. Alexander Horch  
Prof. Dr. Johann Krauser  
Dr. Bruno Schilli  
Dipl.-Inform. Falk Schmidsberger  
Dr. Ulrich Topp  
Dr. Roland Weiss

- Lokale Organisation

Wenke Mulkau  
Dipl.-Inform. Falk Schmidsberger  
Marika Schönebaum



# **Eingeladener Beitrag I**



# Global Parking Space Management: The Case for Database Web Services

Dr. Suo Song, Prof. Dr. Klaus R. Dittrich, Phone: ++41 44 635 4312, Fax: ++41 44 635 6809,  
E-Mail: {congsuo|dittrich}@ifi.unizh.ch, Universität Zürich, Binzmühlestrasse 14, CH-8050 Zürich

## Abstract

As a result of the ever-increasing proliferation of information, a multitude of data sources and software applications are available in world-wide networks that are potentially capable of supporting the efficient development of large scale information systems for various purposes. However, it is still a significant challenge to integrate disparate data sources and other business functionality across organizational and geographical boundaries in a practical manner. Due to their cross-platform interoperability, we believe that web services are a versatile base technology to support global information integration, complementing and often even superseding traditional data integration and enterprise application integration. In this paper, we present our concepts for and experience with leveraging web services to implement inter-enterprise data integration. Based on this integration platform, we have developed a prototype application to provide a European-wide parking space management system that integrates parking services from a variety of parking operators.

We propose to build a customer-specific global view of a set of web services which implement the same or very similar functionalities. This global view enables a service requester to pose a request formulated as an XML document against the global schema (view). This request is referred to as a global request which is then transformed to one or several local requests against the schemas (descriptions) of local web services. Consequently, it offers a simple and pragmatic approach to dynamically access various web services and the resources they expose.

**Keywords:** service-oriented architecture, web services, service composition, data integration, application integration

## 1 Introduction

Modern enterprises are increasingly looking to connect their autonomous information systems with each other to build dynamic, loosely-coupled virtual enterprises, in order to quickly respond to market requirements. By combining resources and processing capabilities of each enterprise, the virtual enterprise can grow much stronger and eventually benefit each participant. Driven by the growing pressure of competition in the market, industry is continuously interested in exploiting appropriate IT infrastructures to effectively construct, maintain, and evolve their information systems and develop novel business applications. Existing enterprise resources and the underlying infrastructure of sharing them furnish plenty of reusable building blocks for composing large-scale, sophisticated e-commerce applications efficiently. While enterprise and inter-enterprise Integration are increasingly crucial to business success, new solutions and approaches are demanded to evolve legacy solutions [7].

The natural combination of the component-based development paradigm (CBD) and the concept of service-oriented architecture (SOA) suggests an elegant approach to build applications by reusing and combining well-designed and thoroughly-tested software

services as building blocks [1][19], enabling enterprises to improve efficiencies throughout their value network and develop greater flexibility in responding to market demands. Although the idea behind this approach roots in the concept of modules and is thus not brand-new, its current significance is the adherence to standards and the ubiquitous support from industry.

The basic essence of SOA is to invoke/use software pieces as services in a standard, secure, and reliable way and combine all participant services together using orchestration/choreography to construct a business application (process) or create a more complex service. The major benefit of SOA lies in the standard ways of exposing and sharing business functionality as loosely-coupled, autonomous, and independent software services that can be invoked by any client that requires the specified functionality. The constituent services within a business application need not be developed by or reside in one company or a single domain of control. The ability to invoke an external service based on standard protocols from another autonomous information system facilitates the development of business applications dramatically. First of all, it introduces a flexible solution of B2B integration. Having a service-oriented architecture and standard specifications and protocols for defining, discovering, and using software services as building blocks simplifies the development of business processes across multiple companies. Second, instead of implementing a specific function itself, an application or a process can invoke an outside service that offers the required function. This simplifies the logic of the overall application since the invoked services remains separated from the application.

While web services are increasingly utilized as a system integration approach, the integration of web services themselves has not been addressed sufficiently. Most current research focuses primarily on processing and using web services individually.

Because web services provide a powerful wrapping mechanism to encapsulate legacy systems, more and more enterprises adopt web services to expose and share their business functionality or data. Even a database system can be wrapped to provide a set of well-defined web services [6]. Therefore, a web service can be regarded as a part of a light-weight enterprise information system. Its data are described by its output documents and queries are regarded as its input documents. The schema of a web service is presented by its description. Based on this perspective, we propose to build a global view of a set of web services which implement the same or very similar functionalities. This global view provides a unified interface for clients to access a variety of web services and their underlying information systems.

This paper is organized as follows. We describe the background and related work in section 2. In section 3, we present the framework of a mediator-oriented SOA model in which a web service mediator presents a global view of a set of sibling services. While in section 3 the descriptions are independent of concrete web service standards, we illustrate the web service mediator using concrete examples in section 4 and 5, respectively. In section 6, we describe the prototype of a global parking space management system based on the concept of web service mediator to show its versatility. Section 7 concludes the paper.

## **2 Background and Related Work**

The database research community has made substantial achievements in integrating data from different and autonomous sources (e.g. [8]). The purpose of a data integration system is to provide a unified view of data from a number of sources by predefining a global mediated schema and semantic mappings between the global schema and the sources containing the

real data. In some sense, a global schema binds the data sources and applications using them very tightly in that both cannot change freely. A data integration system is often regarded as a distributed system where a global schema defined using either the GAV (global-as-view) or LAV (local-as-view) approach acts as a "central" point of access and provides the only "path" from the application to the data source.

At the most fundamental level, each web service must have a machine-processable, common standard description to specify the necessary information, such as message formats, data types, transport protocols, and so on, to enable a requester to transparently make use of that service from another program using a standard message transport. A number of specifications have been proposed to describe web services at different levels of abstraction within various models. The predominant proposal at the lowest level is the Web Services Description Language (WSDL) providing a model and an XML format for describing web services in two parts: an abstract part where a web service is described in terms of the messages it sends and receives, and a concrete part at where the details of accessing a web service are specified [20]. The machine-processable WSDL documents offer service requesters the potential to invoke and discover web services autonomously and automatically with reduced human intervention.

To complement WSDL, various specifications and schemes have been proposed for describing the semantics of web services precisely. Depending on the representation approaches adopted, two types of schemes can be identified: explicit representation and implicit representation [16]. A task can be identified precisely by a unique concept and thus a service can be described by enumerating its tasks. The implicit representation approach does not describe the tasks performed by a service explicitly. Instead, it describes a service by characterizing the transformation (e.g. information exchange, state change) effected by its execution (e.g. [11]). Due to the "web-nature" such as diversity, heterogeneity, autonomy, and large scalability behind web services, the implicit representation approach is more appropriate for describe the semantics of web services ([16]).

The goal of discovery is to find appropriate web services that meet certain functional and non-functional criteria. The discovery process can be performed at design or at run time by a human requester using software tools (discovery services), or by a requester agent (a program). The key to facilitate smarter and more flexible automation of service provision (from the provider perspective) and use (from the requester perspective) is the ability to discover, select and invoke the appropriate services dynamically according to the requirements of requesters and the constraints of providers at run time.

In general, a discovery process involves two steps: find and obtain the descriptions of services, and perform a match between service requirements and descriptions (advertisements). There are two major types of approaches to find and obtain the service descriptions: search-oriented and storage-oriented. The former employs a search engine to collect description wherever it finds them by crawling the internet (e.g. [4]), the latter uses storage to keep and organize the service descriptions submitted by service providers actively (e.g. [17]). The storage of service descriptions can be constructed as a registry, an index, or a peer-to-peer system ([19]). To use a service properly, especially in mission-critical business applications, a trusted business service registry (BSR) is preferred because it can improve the protection of both the service requesters against malicious providers, and the service providers against malicious requesters. There are currently two most eminent business service registries: the Universal Description, Discovery, and Integration (UDDI) specification and the ebXML registry.

The most crucial issue in service discovery is to match service requests with service descriptions to find the most appropriate service. Based on a WSDL/UDDI combination, most early approaches for service discovery are based on keyword matching of WSDL descriptions, and businesses, services, and so-called tModels in UDDI repository. However, to find a desired web service based on conventional keyword search on web service descriptions is far from sufficient because keywords are not able to capture the underlying semantics of web services precisely and completely.

A web service can be used either in an application run directly by humans or in a composite service/application within other programs. There are various patterns, with different complexity, of using web services. The simplest pattern is to invoke a single operation on the specified web service independently. The mechanism of invoking an operation forms the foundation for more complex patterns of using web services: coordination patterns, aggregation patterns and composition patterns.

The SOAP protocol has been proposed to support all interactions among web services across the internet. SOAP specifies messages using XML in a structured and typed manner and defines bindings to actual transport protocols such as HTTP or SMTP ([15]). Almost every software vendor supports SOAP as the common standard of interacting with a web service.

SOAP provides basic interoperability and a web service can be invoked by sending XML documents packaged into SOAP messages. The SOAP protocol basically supports a stateless, one-way message exchange paradigm on which more complex interaction patterns (e.g. request/response, request/multiple responses, solicit/response) can be built.

### **3 Global Views of Web Services**

The essence of a Service Oriented Architecture (SOA) is to (re)use software pieces as services in a standard, secure, and manageable way and to orchestrate them to obtain a variety of business applications (processes) or more complex services. This architectural paradigm combined with standard technologies such as web services greatly facilitates interoperability among heterogeneous systems by establishing a loose coupling between a service provider and a service requester.

Web services target larger and more pragmatic internet-scale interoperability than any previous technology such as CORBA, DCOM or RMI. Interoperability is now based on document-oriented computing rather than distributed object technologies [18]. A web service is described by an XML document (e.g. a WSDL document), is designed to process an XML document (e.g. a SOAP message) as its input request and produce an XML document as its output response. A client (service request) is able to automatically or semi-automatically discover a web service by parsing its description schema and to interact with a web service by composing and sending an XML document.

In a real world application, a client often needs to access a number of web services which implement same or similar functionalities but are owned by disparate providers. Currently, this client has to compose a different XML document for each distinct web service and to handle different XML documents received from different web services.



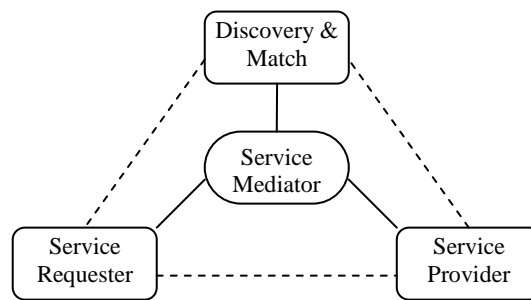


Figure 1. General Mediator-Oriented SOA Framework

Inspired by successes in the domain of data integration and data exchange (e.g. [5][8][9][14]), we propose a mediator-oriented SOA framework to present a unified view of a web service aggregated from a variety of similar web services, referred to as *sibling services*. As depicted in figure 1, the classical SOA model is enhanced by introducing a mediator (or mediators) in the center. A mediator maintains a global schema as the description of a virtual web service representing a number of local web services. The mappings between the global schema and each local schema are defined and used to specify the translation of web service requests and responses. Consequently, a client does not need to compose a distinct request against each web service. Instead, it poses a request against the (virtual) representative web service according to the global schema. The related mediator is responsible to translate this global request to a local request against a local schema. If a web service produces a (local) response, the mediator translates it to a global response which can be processed by the specified client according to the global schema.

The core concepts in this abstract framework are:

- a) a global schema for a number of sibling web services
- b) a set of mappings between a global schema and each local schema
- c) a set of rules of generating and maintaining global schemas and mappings
- d) a set of rules of translating web service requests and responses

We describe the details of this general framework in the following subsections.

### 3.1 Schema Matching and Mapping

We refer to a web service schema (WSS) as a document describing the interfaces to and semantics of a web service [19]. A global schema is defined as the description of a virtual web service representing a finite number of local “actual” web services. There are two fundamental ways of defining such a global schema: *push* and *pull* from the perspective of the mediator.

The first way is to predefine a global schema by a mediator for its representative web service and *push* this description to web service providers or registries to search for similar web services which. Furthermore, a provider can use this global schema as a reference to develop its own web service. It is not mandatory for a provider to implement its web service exactly as the global schema dictates. Instead, the service provider may specify the mappings between its own schema and the general schema. Mappings can be implemented either at the mediator-side or at the local service-side.

The second way is to *pull* descriptions from web service providers or service registries and classify web services into categories or clusters using similarity search (e.g. [4]). After a

number of sibling services are identified, schema matching techniques such as [2][9][14] can be employed to generate a mediated schema and related mappings.

The *push* method is more appropriate for startup scenarios in which only few web services exist. The *pull* method is more appropriate for scenarios in which a large number of web services are already available.

Pragmatically, a hybrid approach may be appropriate. A mediator initiates itself by searching and pulling schemas of web services to construct a global schema to serve a multitude of clients. By learning the requests issued by different clients to a global schema, the mediator is able to adapt itself to serve clients better and pushes this global schema to service providers to make more providers' web services available.

Once the global schema is defined, various approaches [14] can be combined to produce a mapping set between the global schema and each local schema. The overall process is depicted in figure 2. A minimal description of a web service is formulated as **WSS**: ( $\mathbf{OP}$ ,  $\mathbf{I}^{\mathbf{OP}}$ ,  $\mathbf{O}^{\mathbf{OP}}$ ), where  $\mathbf{OP}$  is a set of operations defined for a web service,  $\mathbf{I}^{\mathbf{OP}}$  is the set of input messages (requests) of operations, and  $\mathbf{O}^{\mathbf{OP}}$  is the set of output messages (responses). Consequently, the web service schema matching and mapping can be defined as ( $\mathbf{G}$ ,  $\mathbf{L}$ ,  $\mathbf{GL}$ ,  $\mathbf{LG}$ ), where  $\mathbf{G}$  is the global schema,  $\mathbf{L}$  is the set of local schemas,  $\mathbf{GL}$  is the mapping from the global schema to local schemas, and  $\mathbf{LG}$  is the mapping from the local schemas to the global schema.

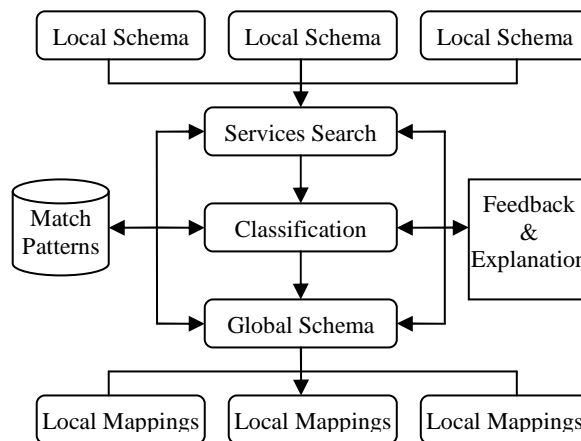


Figure 2. Schema Matching and Mapping

A local schema  $L_j$  is called *conformable* to a global schema  $G$  if each operation  $OP^G$  defined in global schema  $G$  has one conformable operation  $OP^{L_j}$  defined in schema  $L_j$ . In turn, an operation  $OP^S$  defined in schema  $S$  is called conformable to another operation  $OP^T$  defined in schema  $T$  if

- the semantics of the functionality of  $OP^S$  is identical to  $OP^T$
- the  $OP^S$  input message can be composed by manipulating the  $OP^T$  input message
- the  $OP^S$  output message can be composed by manipulating the  $OP^T$  output message

The primitive elements in a web service schema which are meaningful to schema matching are web service operations. Since web service operations can be grouped into different web services, a global schema may have different matching level: *operation-level*, *service-level* (a service consists of multiple operations), and *process-level* (a process consists of multiple services).

Given a global schema  $G$ , the problem is to discover and match as many as possible conformable local web services. Given a set of sibling services, we have to determine whether a global schema for them exists, or compute one if not. If a global schema cannot be derived for a set of sibling services, a subset of the for which a global schema can be defined is computed. Those local services which caused the creation of a global schema to fail are listed and the reasons are presented to the mediator.

Ideally, matching and mapping are processed in a fully automated way. However, the state-of-the-art of schema matching systems shows that this is largely performed manually [2][14]. In most real applications, the matching between some elements in different schemas cannot be defined in a precise mathematical expression which would be crucial to the design of match algorithms [14].

One major bottleneck is the determination of the semantics of the functionality of web service operations. Although a lot of research efforts have been made, human intervention is still required in some cases to decide whether two operations perform the same functionality. A further bottleneck is the intrinsic complexity of the matching structure between schemas. One element in a schema could match many elements in another one. Especially, when custom data types are used, more complex matches are required. Currently, the most pragmatic solutions are semi-automated (e.g. [2][9]) and discover matches automatically according to some general or domain-specific principles, and human intervention is required to confirm the valid matches and correct the invalid ones.

### 3.2 Request/Response Transformation

A client sends a web service request composed according to the global schema. This request, however, cannot be processed directly by an actual local web service. Therefore, the mediator has to translate the global request to a specified local request. Similarly, a local web service produces a response document which has to be translated by the mediator for the actual service requester.

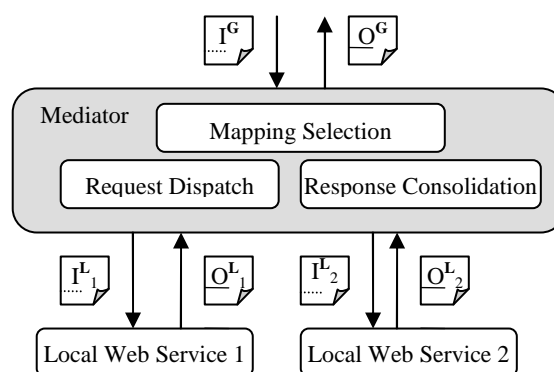


Figure 3. Request/Response Translation

When a mediator receives a service request from a client, the first task is to determine and select local web services to process this incoming request. This selection can be specified either explicitly by the client in the request or implicitly according to the constraints specified by the client. A service request can be served by multiple services, for example, to retrieve flight information from different airline information systems.

When the local services are selected, the mediator rewrites the original incoming request for each selected local service using related mappings  $GL$ . As shown in figure 3, an incoming request  $I^G$  is transformed into two local requests  $I^{L_1}$  and  $I^{L_2}$  which are delivered to corresponding local web services respectively.

The local web service processes the incoming request  $I^{L_1}$  ( $I^{L_2}$ ) and produces a response document  $O^{L_1}$  ( $O^{L_2}$ ) if required. The response documents are then sent back to the mediator which can either simply merge the two documents or consolidate them (e.g. remove duplicate results) and transform the consolidated document to the global response document  $O^G$ .

### 3.3 Match Patterns

In order to facilitate the automation of schema matching, domain knowledge and external information are required to help the determination of match candidates [2][14]. We organize such knowledge and information as *match patterns* which can be used in schema matching and request/response transformation.

A major part of the match patterns is organized to define the relationships between common data types or domain-specific data types. For example,

- document element name match patterns:  
Last Name ~ Family Name  
Name ~ (First Name, Middle Initial, Last Name)  
Address (Street, City, ZIP, Country) ~ Location (Street, Town, Code, Country)
- document element value match patterns:  
Reservation (from, to) ~ Reservation (from, duration) || duration = to – from  
TotalPrice = ReservationFee + ParkingFee
- data type match patterns:  
Array [3] ~ String1, String2, String3

Data type match patterns are often applied together with name match patterns. They are useful in cases where web service developers define their own custom data types, for example, if a web service returns flight information as an array and another web service returns flight information as a string delimited by commas. Then the mediator of the airline information service should be able to convert between both cases.

### 3.4 Mediator Implementation

Not surprisingly, the web service mediator itself is designed as a web service that is able to process incoming requests formulated as XML documents, and produce output XML documents as responses. Considering the web services architectural model [19], a mediator can be implemented in two different ways.

The first approach is to implement the mediator as a *web service intermediary*, which is defined as a web service to transform documents (requests/responses) [19]. A mediator is located in the middle of the message path between the original document sender and the ultimate receiver. The client sends a document while it can ask the intermediary to modify the document in order to match the request schema of the ultimate receiver. However, this

requires some extension mechanisms which may cause interoperability problems because of different implementations of web services infrastructures.

The second approach is to implement the mediator as an independent web service. The client composes an XML document as request according to the mediator's global schema and expects an analogous response. From the perspective of the client, the ultimate receiver of its request is the mediator, not the actual service performing the requested functionality.

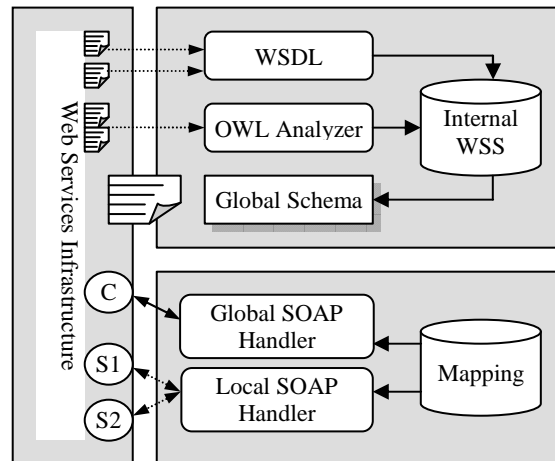


Figure 4. The Web Service Mediator Architecture

As shown in figure 4, we have implemented a web service mediator as an independent web service based on standard web services infrastructure: we describe and discover web services using WSDL documents, and call web services by sending and receiving SOAP messages. A web service mediator is composed by a matcher that performs schema matching and mapping and a caller that performs request/response transformation. The matcher component consists of several schema-specific analyzers (such as a WSDL analyzer) which convey schemas of web services into an internal concise expression. As depicted in figure 2, a global schema is generated and published so a client can interact with any local web service by composing requests according to a uniform schema. An incoming global SOAP request is transformed to one or several local requests and re-sent using the web service infrastructure. The local responses are collected and consolidated by the global handler to one global SOAP response which is sent back to the original requester.

## 4 WSDL Schema Matching

Because WSDL is the current de facto standard of describing production web services, we first focus on implementing a WSDL analyzer to convert a WSDL document into the internal WSS format.

A WSDL document defines a web service with an abstract description in terms of the messages it sends and receives, and a concrete description in terms of the transport and wire format details for its supported interfaces. When the WSDL analyzer receives a WSDL document, the concrete bindings of this service are retrieved and associated with an internal unique ID of that web service. This information is later used for interacting with this web service.

The more important part of the WSDL document is the matching of the in/out messages against the messages defined by the global schema. As an example, the piece of a WSDL

document in figure 5 shows the abstract information of a car park reservation web service. For simplicity, some details have been. A web service belonging to CarPark1 has offered a web service operation reservePS which accepts an XML document with elements (driver, from, to, lot) as input request to reserve a parking space at a time specified in the request.

```
<types>
  <s:element name="reservePS">
    <s:complexType>
      <s:sequence>
        <s:element name="driver" type="s:string" />
        <s:element name="from" type="s:time" />
        <s:element name="to" type="s:time" />
        <s:element name="lot" type="s:long" />
      </s:sequence>
    </s:complexType>
  </types>
  <message name="reservePSSoapIn">
    <part name="parameters" element="s0:reservePS"/>
  </message>
  <portType name="CarPark1ReservationSoap">
    <operation name="reservePS">
      <input message="s0:reservePSSoapIn"/>
      <output message="s0:reservePSSoapOut"/>
    </operation>
  </portType>
```

Figure 5. A local web service schema

A global schema of a generic web service is presented in figure 6. The defined operation `MakeReservation` accepts an XML document of elements (`cpID`, `driver`, `from`, `to`, `amount`) as input request. Although the example is quite simple, it is sufficient to illustrate the basic concepts of WSDL document matching.

To determine whether an operation `reservePS` is conformable to a global operation `MakeReservation`, the following tasks have to be performed:

- Semantics of functionality: Since there is no standard way to convey such information within WSDL documents, the final determination is made semi-automatically assisted by the associated match patterns and a domain-specific ontology library.
- Input message matching. The principle of input message matching is simple, i.e., it just has to be checked whether it is possible to compose a local input message using the data from the global input message. In the global input message, the element `cpID` is consumed by the mediator itself to select the appropriate web service. In this case, the client specifies the owner of the local web service explicitly.
- The global element `parker` is mapped directly to the local element `driver`;
- The global element `entranceTime` is mapped directly to the local element `from`;
- The global element `duration` is mapped to the local element `to` using the match pattern `to = entranceTime + duration`;

- The global element `amount` does not have a matching local element which means it will be ignored. Semantically, this means the local reservation does not allow one request for multiple parking spaces;
- Finally, the local element `lot` cannot be generated from the global request data. Semantically, this means that the local reservation service allows the driver to specify where to park her car, but the global service does not support this feature. If there are no other business rules defined as match pattern, the match result is that the local service `reservePS` is not conformable to the global service `MakeReservation`.

```
<types>
  <s:element name="MakeReservation">
    <s:complexType>
      <s:sequence>
        <s:element name="cpID" type="s:string" />
        <s:element name="parker" type="s:string" />
        <s:element name="entranceTime" type="s:time" />
        <s:element name="duration" type="s:time" />
        <s:element name="amount" type="s:long" />
      </s:sequence>
    </s:complexType>
  </types>
  <message name="MakeReservationSoapIn">
    <part name="parameters"
      element="s0:MakeReservation"/>
  </message>
  <portType name="GenericReservationSoap">
    <operation name="MakeReservation">
      <input message="s0:MakeReservationSoapIn"/>
      <output message="s0:MakeReservationSoapOut"/>
    </operation>
  </portType>
```

Figure 6. A global web service schema

#### 4.1 Handling Missing Data

Values of elements in a local schema which cannot be acquired from the data of a global request is referred to as missing data. There are three ways to handle missing data to improve the accessibility of local web services.

The first approach is to expand the global service schema to include all elements of every local schema. However, this is not always appropriate, and will result in a huge global request carrying many elements where most of them will be ignored by most local services. The second approach is to declare an element as optional. For example, if the element `lot` is optional, the mediator can just leave it empty or set it to an empty string. The local web service will then just ignore this element. Therefore, if a match pattern is added to state that `lot` is optional for service operation `reservePS`, the match result is changed. The third approach is to store some additional application-specific information about clients within the mediator, which can then provide a concise interface because some elements can be composed using the stored information. For example, in our prototype, we store `parker` information within the mediator. There is thus no need for the client to compose the request with more elements such as `parker` name, address, and other identity data each time. Clearly, this approach is a design decision based on specific application requirements and constraints.

## 4.2 WSDL Versions

The purpose of our prototype is to present a unified view of parking services from disparate operators. Most parking operators within our project have published their web services using WSDL 1.1. Therefore, we have built a WSDL analyzer for WSDL 1.1 documents.

In addition, according to the migrating concerns in the WSDL 2.0 specification [20], the following mappings are required between a WSDL 1.1 document and a WSDL 2.0 document:

- PortType (1.1) ~ Interface (2.0)
- Port (1.1) ~ Endpoint (2.0)

## 4.3 Interaction Mode and Encoding Style

Web services accept XML documents as requests and produce XML documents as responses. A client interacts with a web service by composing the requests and consuming the responses. Typically, the low level processing of XML documents is performed by the underlying web services infrastructure where two interaction modes (binding styles) of higher levels are defined in WSDL documents: RPC-style and document-style interaction.

The WSDL document can specify the encoding style which defines how data is packaged and sent over the wire. Therefore, there are four combinations of interaction mode and data encoding.

In the following section, we illustrate how to transform SOAP messages using RPC-literal and document-literal styles.

# 5 SOAP Message Transformation

SOAP is a specification for interactions among web services across the Internet. SOAP uses XML to exchange structured and typed information and defines bindings to actual transport protocols such as HTTP or SMTP [15]. Most software vendors support SOAP as the common specification for interacting with a web service. Consequently, we have implemented a SOAP handler in the web service mediator to transform SOAP messages between global requests/responses and local request/responses.

SOAP supports both RPC-style and document-style interaction modes. The major difference is how service requests and service responses are handled by the service provider and the service requester based on the underlying web service infrastructure.

In the following subsections, we illustrate how a SOAP message is transformed into another structure with different types. Within our prototype implementation, we only consider SOAP version 1.2. The transformation between SOAP 1.1 document and SOAP 1.2 document has not been addressed.

## 5.1 RPC-Style SOAP Messages

Within the RPC-style SOAP paradigm, SOAP messages are translated into corresponding RPC method signatures, i.e. method names and input/output parameters. The transformation of an RPC-style message into another RPC-style message can be performed at two levels: RPC method level and SOAP message level.



The transformation at RPC method level works as follows:

- A client calls a global RPC method;
- When this global method is invoked, it receives the input parameters from the request via the underlying web service infrastructure;
- The global method maps the input parameters (global) to input parameters of a local method to be invoked;
- The global method calls the local method and receives the return value;
- The global method maps the return value into its own return value passed on to the original client.

The advantage of this approach is that it utilizes the underlying web service infrastructure to process the SOAP messages and the mappings are performed between method parameters. Compared with the RPC method level transformation, the message level transformation has its own advantages, in particular with respect to performance and interoperability. We illustrate the message-level transformation in the next section using document-style SOAP messages.

## 5.2 Document-Style SOAP Messages

Compared with RPC-style, the document-style SOAP paradigm is more loosely coupled because the major processing of the XML documents as requests and responses are performed at the application level (i.e. by the requester and provider) rather than at the infrastructure level (i.e. by the proxy code passing the parameters).

Suppose we have two local parking services which enable a client to enquire which car parks have available parking spaces. The Java interfaces are presented in figure 7.

```
Local Service A
...
public String [] getCarPark() {
...
}

Local Service B
...
public String getCarParkList() {
...
}

The global service is defined as:
...
public String [] getAvailableCarPark() {
...
}
```

Figure 7. Web Services Interfaces

As shown in figure 7, there are two local service operations. The first one accepts no parameters and returns an array of strings of the names of car parks which have empty spaces during the demanded time. The second operation accepts no parameters and returns

a comma-delimited string of names of car parks. The global service operation `getAvailableCarPark` also takes no parameters, meaning that it will automatically select local web services. Without any further constraints, this global service will invoke all local web services in its matching set. The response SOAP messages are presented in figure 8 and figure 9 respectively.

```
<soap-env: body>
  <ns1:getCarParkResponse>
    <return xsi:type="ns2:array"
      ns2:arrayType="xsd:string[5]">
      <item xsi:type="xsd:string">CP1</item>
      <item xsi:type="xsd:string">CP2</item>
      <item xsi:type="xsd:string">CP3</item>
      <item xsi:type="xsd:string">CP4</item>
      <item xsi:type="xsd:string">CP5</item>
    </return>
  </ns1:getCarParkResponse>
</soap-env:body>
```

**Figure 8. Response from Local Service A**

The same car park CP2 has appeared in both the returned result of local service operation `getCarPark` and in the returned result of local service operation `getCarParkList`. Therefore, the global operation `getAvailableCarPark` receives the XML documents shown in figure 8/9, and produces a consolidated return message sent to the original service requester as shown in figure 10.

```
<soap-env: body>
  <ns1:getCarParkListResponse>
    <result xsi:type="xsd:string" >
      CP2, CP7, CP9
    </result>
  </ns1:getCarParkListResponse>
</soap-env:body>
```

**Figure 9. Response from Local Service B**

```
<soap-env: body>
  <ns1:getAvailableCarParkResponse>
    <return xsi:type="ns2:array"
      ns2:arrayType="xsd:string[5]">
      <item xsi:type="xsd:string">CP1</item>
      <item xsi:type="xsd:string">CP2</item>
      <item xsi:type="xsd:string">CP3</item>
      <item xsi:type="xsd:string">CP4</item>
      <item xsi:type="xsd:string">CP5</item>
      <item xsi:type="xsd:string">CP7</item>
      <item xsi:type="xsd:string">CP9</item>
    </return>
  </ns1:getAvailableCarParkResponse>
</soap-env:body>
```

**Figure 10. Response from Local Service A**

## 6 Proof of Concept

In this section, we describe a general parking space management application based on the concept of building a global view of a set of sibling services which implement similar functionalities. We have already alluded to some details this example application above.

### 6.1 Requirements

Parking resources and services have significant influence upon the efforts towards enhancing urban mobility and improving the usage of urban space. In most European cities, parking space is a very limited resource and the geographical distribution of parking spaces is often not proportional to the actual parking requirements of the public. A lot of academic research efforts and industrial developments have been contributed to build innovative information systems which disseminate parking information, such as when, where, and which parking space is available, to the drivers, in various ways. With such information, drivers can be guided to find and reserve the required parking spaces easily. Thus parking space search activities as well as the corresponding traffic can be decreased remarkably.

Nowadays, more and more parking space providers are increasingly interested in utilizing online information systems to facilitate customers to find, reserve, and use parking spaces easily. In order to optimize the use of these information systems, the providers need to offer their online services in a way that any potential client (a human user or a software agent on behalf of a human user) can find and access the desired service flexibly via various network-enabled devices. In essence, providers want their services to reach as many customers as possible and customers (service requesters) want to find desired services easily on demand.

However, due to the intrinsic heterogeneity among parking space providers, users are facing many information silos: each provider develops its own information system independently and thus often serves the customers in a different way than others. It is very inconvenient for a human user to find/reserve a suitable parking space through these information silos if she is not familiar with car parks around her destination. Without a well-structured collection of related parking information from all car parks, an automatic accurate search for parking spaces is almost impossible. While these parking services might be consumable in a limited way by human users who crawl through the web to find the access point of a desired service from a multitude of distinct web pages, it is very difficult to access these services from other business applications (e.g. for making reservations at hotels, restaurants, festival events) which represent a large amount of indirect users of parking services. An undesirable consequence is that the heterogeneity hinders potential users/drivers from reaching parking space providers.

Due to the existence of service heterogeneity, a piece of specific custom code to handle the use of each native service is required. Thus a design strategy must be considered carefully before realizing the functionality of searching and booking parking spaces at any desired car park according to a customer's requirements. If there are only a few car parks, it is certainly feasible to implement the logic of service selection and invocation in the client application directly. However the application will quickly become cumbersome when the number of parking services increases, resulting in applications that are hard to manage and evolve; it is not a scalable solution. The client application needs to introduce new custom code every time when a new parking service is published. Furthermore, state-of-the-art service discovery mechanisms are still not able to discover a service matching a specific requirement precisely. It is not conceivable to use a "sufficiently similar" service in business applications without human confirmation. Thus, each service to be used in business applications must be verified with human assistance. If the logic of handling the use of each service is completely

implemented inside the client applications, their developers must have the knowledge of how to use all services involved. When a new service is published, each application that needs to use it has to be updated to include a piece of custom code handling the new service.

## 6.2 iPARK: Universal Parking Services

When a particular application needs to use a large number of services realizing the same functionality and are available from various providers, it is a rational strategy to separate some parts of the logic of selecting and using the services from the application. This vision leads to the concept of service mediator which presents a unified view of a (virtual) global service representing a number of conformable local web services.

The principal innovation behind our iPARK model (figure 11) is the idea of building and operating a dynamic system which provides universal parking services and hides the differences in parking services offered by each local parking management service (PMS) system.

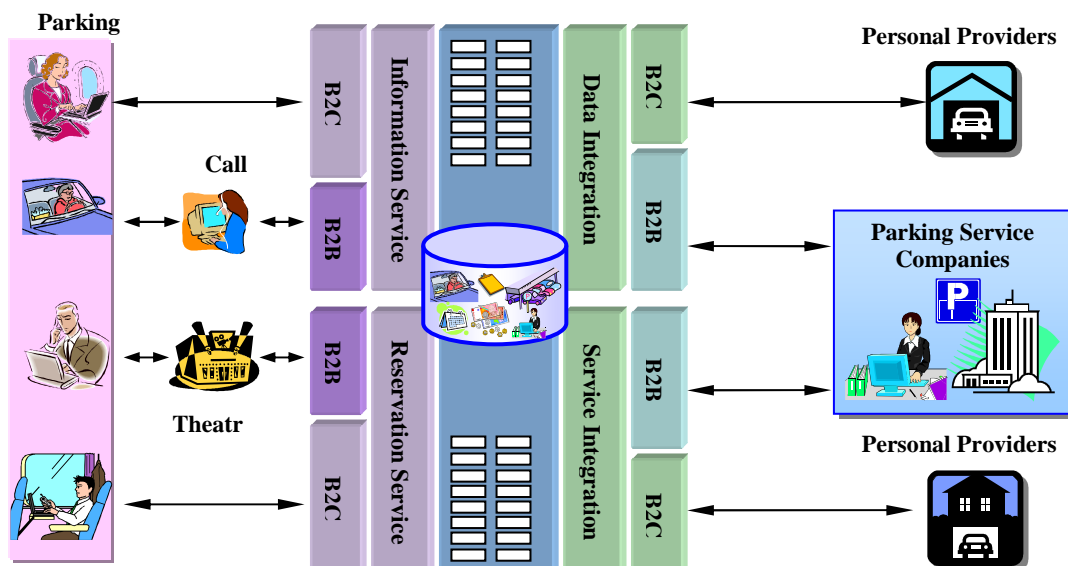


Figure 11: The iPARK System

An important feature of iPARK is the creation of a *data service tier* which achieves three goals: 1) a clean data tier where the code and data are still separated; 2) a mediation tier where the code and data are combined by wrapping access to the underlying database with a set of web services; 3) the separation of the actual underlying database schema and several public virtual schemas defined in terms of web services.

The data service tier is the kernel of the virtual business layer, connecting data and business logic. Leveraging web services to encapsulate data manipulation, the iPARK application achieves high interoperability and scalability.

Because the actual business logic accesses data via a set of web services, a loose coupling between the database and the business logic is obtained. Therefore, privacy is protected because the underlying database schema can be hidden from the (external) business logic.

The success of traditional data integration approaches is limited for enterprise applications because the heterogeneity of underlying information systems impedes access to data across enterprise boundaries. Complementing traditional data integration, the use of web services to wrap a database facilitates inter-enterprise data integration.

Based on current technologies, we believe that at this stage, the best way to connect an enterprise database to internet-based commercial applications is to wrap it with a set of web services which defines a public view of the underlying database. The enterprise can define different policies of wrapping the database with different sets of web services and then expose different sets of web services to different partners.

The iPARK system evolved from an earlier prototype of ours, PSOS. When developing the PSOS system, we made extensive use of an object-relational database by creating a number of data access methods. On the one hand, this facilitated the development of complex database-centered applications. However, on the other hand, it tied the applications tightly to the database. To overcome this limitation, we completely re-engineered the iPARK architecture. All data access methods have been removed from the data tier and a new tier was introduced between the data and the business logic requiring the data.

For this new data service tier, a set of web services have been defined. Although there are several good candidates for describing web services (such as OWL-S), we use WSDL for pragmatic reasons. A set of custom message types have been defined as a public schema of the underlying database. Each web service operation is annotated with an input message type and/or an output message type.

One primary objective of the iPARK system is to provide one consistent user interface to human customers including both parking consumers for accessing parking services (resources) and parking operators for providing parking services. In principal, the iPARK front-end applications provide a parking service portal for end users. The big value to end users of the iPARK system depends largely on the number of parking operators who connect their parking management systems to it. The more operators join iPARK's parking service federation, the larger benefit will result for consumers looking for parking services. With connections to a large variety of parking operators, it provides a powerful parking services portal to end users based on a universal parking service platform built around the iPARK system as the kernel and the local PMS systems as the peripherals. This requires the iPARK system to support a flexible and scaleable way to implement the interactions with peripheral IT systems of parking operators. In order to simplify and encourage parking operators to connect their PMS systems to iPARK, the integration process has to be simple and automated as much as possible with little human efforts.

An automated or semi-automated process is required to bridge the abstract iPARK global service and the actual parking services. Semantic technologies provide facilities to improve automation. However, some manual efforts are still needed to 1) predefine the common semantics of data, messages, and operations; 2) map differences which cannot be handled automatically by programs. Obviously, open standards for parking services and the PMS's conformance to them are vital for the automated interoperability and reduce manual efforts.

The iPARK prototype has demonstrated successfully the ability to create and maintain dynamic connections to 'dumb' parking services which are not actual production web services. However, a full-fledged iPARK system requires parking operators to update their parking infrastructures accordingly.

## **7 Conclusion**

We have presented a novel architectural pattern to extend the classical SOA model with a service mediator. A service mediator presents a unified view, described by means of WSDL, of a global web service that represents a set of sibling web services implementing similar business functionality based on similar XML documents as requests and/or responses.

A web service is described by a schema as a WSDL document (or an OWL-s document). A set of sibling web services can share a global schema which can be used to map the XML-based web service requests/responses. Therefore, a service requester which needs to access a number of sibling services can be freed from low level repetitive logic of handling the request/response transformations. This mapping and matching logic can be separated from the application-level and moved to a mediator between an actual service provider and a service requester. The performance penalty caused by the extra message processing needed is compensated by improved interoperability and simplified application development.

We have not addressed security issues and transaction processing in the web service mediator systematically so far, but plan to do so in our next steps.

## References

- [1] Alonso, G., Casati, F., Kuno, H., and Machiraju, V. *Web Services: Concepts, Architectures and Applications*. Springer-Verlag, 2004.
- [2] Dhamankar, R., Lee, R., Doan, A., and et al. iMAP: Discovering Complex Semantic Matches between Database Schemas. In *Proc. of the SIGMOD 2004*.
- [3] Dogac, A. Kabak, Y., Laleci, G., and et al. Semantically Enriched Web Services for the Travel Industry. *SIGMOD Record*, 33(3), Sep. 2004.
- [4] Dong, X., Halevy, A., Madhavan, J., and et al. Similarity Search for Web Services. In *Proc. of the VLDB 2004*.
- [5] Fagin, R., Kolaitis, P.G., and Popa, L. Data Exchange: Getting to the Core. *ACM Trans. Database*, 30(1), Mar. 2005, 174-210.
- [6] Gray, J. The Next Database Revolution. In *Proc. of the SIGMOD 2004*.
- [7] Halevy, A., Ashish, N., Bitton, D., and et al. Enterprise Information Integration: Successes, Challenges, and Controversies. In *Proc. of the SIGMOD 2005*.
- [8] Lenzerini, M. Data Integration: A Theoretical Perspective. In *Proc. of the PODS 2002*.
- [9] Miller, R. J., Haas, L.M., Hernandez, M. Schema mapping as query discovery. In *Proc. of VLDB 2000*.
- [10] Milo, T., Abiteboul, S., Amann, B., and et al. Exchanging Intensional XML Data. In *Proc. of the SIGMOD 2003*.
- [11] OWL Service Coalition. OWL-S: Semantic Markup for Web Services. <http://www.daml.org/services/owl-s/1.1B/owl-s.pdf>
- [12] Paolucci, M., Kawamura, T., Payne, T. R. and Sycara. K. Semantic Matching of Web Services Capabilities. In *Proc. of the ISWC 2002*.
- [13] Park, J. and Ram, S. Information Systems Interoperability: What Lies Beneath? *ACM Trans. Information Systems*, 22(4), Oct. 2004, 595-632.
- [14] Rahm, E. and Bernstein, P.A. A survey of approaches to automatic schema matching. *VLDB Journal* 10, 2001.
- [15] SOAP. SOAP Version 1.2. <http://www.w3.org/TR/soap12>.
- [16] Sycara, K., Paolucci, M., Ankolekar, A., and Srinivasan, N. Automated Discovery, Interaction and Composition of Semantic Web Services. *Journal of Web Semantics*, 1(1), September 2003.
- [17] UDDI. <http://uddi.org/>.
- [18] Vogels, W. Web Services Are Not Distributed Objects. *IEEE Internet Computing*, Nov. Dec., 2003.
- [19] Web Services Architecture. <http://www.w3.org/TR/ws-arch/>.
- [20] WSDL. <http://www.w3.org/TR/wsdl20/>.
- [21] Wiederhold, G. Mediators in the Architecture of Future Information Systems. *The IEEE Computer* 25(3), 1992.





# **IT-Sicherheit und E-Government**



## Automatisierte Kopieerkennung am Beispiel Eintrittsticket

Roland Krüger, Phone: ++49.201-43752-70, Fax: ++49.201-43752-77,  
E Mail: rKrueger@MediaSec.de, MediaSec Technologies GmbH, Berliner Platz 6-8, 45127 Essen

Vertr. Prof. Dr. Ursula Oesing, Phone: ++49.201-43752-48, Fax: ++49.201-43752-77,  
E Mail: Ursula.Oesing@FH-Duesseldorf.de, FH Düsseldorf, Josef-Gockeln-Str. 9, 40474 Düsseldorf

Dr. Jan C. Vorbrüggen, Phone: ++49.201-43752-70, Fax: ++49.201-43752-77,  
E Mail: jVorbrueggen@MediaSec.de, MediaSec Technologies GmbH, Berliner Platz 6-8, 45127 Essen

Carlos Reuthe, Phone: ++49.160-96444766, Fax: ++49.201-49.43752-77,  
E Mail: u15702@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

### Abstract

Die Anforderungen an die Sicherheit nehmen bei internationalen Großveranstaltungen immer mehr zu. Insbesondere will man den Zutritt zu Veranstaltungen und Veranstaltungsbereichen kontrollieren. Alle Personen mit besonderen Berechtigungen (Presse, Catering, Service, Sicherheitsdienst etc.) werden vorab registriert. Anschließend werden diesen Personen aufgabenbezogene Berechtigungen erteilt. Diese Zutrittsberechtigungen müssen an den Übergangsstellen zu den einzelnen Veranstaltungsbereichen kontrolliert werden können. Dazu muss die entsprechende Person zunächst identifiziert werden, um anschließend ihre Zutrittsberechtigung zu kontrollieren. In diesem Arbeitspapier werden die Anforderungen an mobile Zutrittskontrollsysteme vorgestellt. Auf der Basis der identifizierten Anforderungen wird ein Lösungskonzept für papierbasierte Zutrittskontrollkarten (kurz: Tickets) vorgeschlagen. Die Vorteile von papierbasierten Lösungen sind die geringen Kosten. Die Nachteile papierbasierter Tickets sind der fehlende Manipulationsschutz und die fehlende technische Überprüfbarkeit solcher Tickets auf Integrität und Originalität. Um diese Nachteile zu beheben und gleichzeitig an dem kostengünstigen Papierausdruck festhalten zu können, werden diverse Sicherheitsmechanismen eingesetzt, die auf elektronischen Signaturen, zweidimensionaler Barcodetechnologie und einem sogenannten „Copy Detection Pattern“ basieren.

**Keywords:** Copy Detection Pattern (CDP), Kopieerkennung, Manipulationsschutz, Verpackungsschutz, Sichere Tickets, Originalität

## 1 Einführung

Der Organisationsaufwand bei Großveranstaltungen nimmt mehr und mehr zu. Hierbei stellt die Zutrittskontrolle eine wesentliche Aufgabe dar. Die Zutrittskontrolle hatte ursprünglich nur die Aufgabe, die Entrichtung eines entsprechenden Eintrittsentgeltes sicherzustellen und ggf. eine Personenkontrolle auf unerlaubte Gegenstände durchführen zu können. Heutzutage stellt sich die Aufgabe der Zutrittskontrolle aber sehr viel differenzierter dar. Im folgenden wird auf das Beispiel einer Fußballgroßveranstaltung mit unterschiedlichen Spielstätten eingegangen. Bei solchen Veranstaltungen ist zwischen einzelnen Personengruppen zu unterscheiden. So gibt es beispielsweise die einzelnen Mannschaften mit den zugehörigen Betreuern, Schiedsrichter, Funktionäre, lokale Organisatoren, offizielle Vertreter der Verbände, Reporter und das Servicepersonal. Diese Personengruppen besitzen jeweils unterschiedliche Berechtigungen beim Zugang zu den einzelnen Bereichen, wobei die Zutrittsberechtigungen temporär verschieden sein können und es zudem unterschiedliche Spielstätten gibt.

Es sei hier angemerkt, dass in der Vergangenheit keinerlei technisch unterstützte Kontrollen durchgeführt wurden. Die Erfahrung bei ähnlichen Veranstaltungen hat gezeigt, dass man mit selbstbewusstem Auftreten jede Kontrolle dieser Art penetrieren konnte, auch ohne über

die eigentlich erforderliche Berechtigung zu verfügen. Außerdem konnte vor allem bei den Dienstleistern nicht erreicht werden, dass vom insgesamt eingesetzten Personal bei jedem konkreten Spiel nur das tatsächlich eingeteilte Personal einen Zugang erhält.

## 2 Anforderungen

Die besondere Situation bei einer solchen Großveranstaltung ist, dass an den einzelnen Austragungsorten zwar Zutrittskontrollverfahren etabliert sind, diese aber untereinander nicht kompatibel sind. Das Organisationskomitee vergibt die Zutrittsberechtigungen (sogenannte Akkreditierungen) nach Prüfung der Identität und der Gruppenzugehörigkeit zentral<sup>1</sup>. Die Tickets mit den Akkreditierungen können nicht für jedes Stadion individuell erstellt werden, so dass beispielsweise ein Reporter mehrere unterschiedliche Tickets erhält. Die Systeme zur Erstellung solcher stadionspezifischen Tickets sind beim Organisationskomitee gar nicht vorhanden, insbesondere sind nicht an allen Übergängen zu den einzelnen Stadionbereichen Zutrittskontrollsysteme installiert. Die Sicherheit eines Zutrittskontrollsystems ist daher von der praktischen Umsetzungsmöglichkeit abhängig.

Das Kontrollpersonal an den Bereichsübergängen eines Stadions muss die Entscheidung, ob einer Person der Zugang zu einem speziellen Bereich gewährt wird, schnell und sicher treffen können. Langwierige Betrachtungen des Tickets oder gar Rückrufe sind nicht praktikabel. Daher muss unmittelbar und ohne technische Expertise oder Lernphasen entschieden werden können, ob es sich um ein „echtes“ Ticket für diesen Bereich handelt. Eine technisch unterstützte Prüfung ist hier geeignet, wenn sie mobil einsetzbar und flexibel ist und ein eindeutiges Ergebnis liefert (0-1-Entscheidung). Da sich im Verlauf einer solchen Veranstaltung die Zutrittsberechtigungen ändern können, müssen Akkreditierungen auch zurückgezogen werden können. Ein trauriges Beispiel ist das Ausscheiden einer Mannschaft, deren Spieler und Begleitpersonal somit die Akkreditierungen für das nächste Ausscheidungsspiel verlieren.

### Zusammenfassung der Anforderungen an das Ticket

- Dezentrale Erzeugung
- Kostengünstige Erzeugung
- Unterstützung von Standardverfahren
- Ticket trägt personenindividuelle Merkmale zur Identifizierung
- Ein Ticket für alle Veranstaltungen
  - Robustheit des Tickets
- Schutz der Datenintegrität
- Schutz der Originalität
- Mobile und kostengünstige Verifikationsmöglichkeit in einem Schritt
  - Integrität
  - Originalität
- Eindeutige und sichere Entscheidung bei der Verifikation
  - Keine Expertise des Kontrollpersonals notwendig
  - Schnelle Verifikation und einfache Handhabung

---

<sup>1</sup> Aktueller, erkannter Fälschungsfall vom 13.01.2006: <http://www.wdr.de/radio/wdr2/westzeit/322646.phtml>

- Möglichkeit des Zurückziehens der Akkreditierung nach der Ausgabe des Tickets

### 3 Die Umsetzung

Da für die zentrale Erzeugung der Tickets aus Kostengründen nur die übliche Bürohardware zur Verfügung steht, werden Papiertickets erzeugt, die anschließend laminiert werden, um die notwendige Robustheit und Haltbarkeit zu garantieren. Zur personenindividuellen Identifizierung wird neben dem Namen, dem Arbeitgeber und der Gruppenzugehörigkeit ein Lichtbild der zutrittsberechtigten Person auf das Ticket aufgedruckt. Die Prüfung der Übereinstimmung von Lichtbild und der Eintritt begehrenden Person wird zur Hauptprüfaufgabe des Kontrollpersonals, weil die Prüfung der Zutrittsrechte und der Originalität des Tickets vom Kontrollgerät übernommen wird. Zudem werden die „groben“ Zutrittsberechtigungen (Stadion, Spieltag, Seriennummer) menschenlesbar auf das Ticket aufgedruckt.

Für Dienstleister wurde als Innovation vorgesehen, dass zum einen jeder Mitarbeiter sein persönliches Ticket (mit Lichtbild) erhält, das aber allein den Zugang nicht erlaubt. Der Dienstleister als Organisation erhält zusätzlich eine kleinere Anzahl von reinen Zugangstickets, die in Kombination mit dem persönlichen Ticket und übereinstimmendem Organisations-Feld den Zutritt erlauben.

#### 3.1 Integritätsschutz

Die individuellen Zutrittsberechtigungen werden elektronisch signiert und incl. Signatur maschinenlesbar mittels Barcodetechnologie auf das Ticket aufgedruckt. Dieses Verfahren ist kostengünstig und sichert die Integrität der Daten, da der 2D-Barcode leicht mit einem Handheld ausgelesen und ausgewertet werden kann.



Data Matrix



Aztec



RSS-14



QR Code

Abb. 1: Beispiele für 2-dimensionale Barcodes

#### 3.2 Schutz der Originalität

Desweiteren wird zum Schutz der Originalität ein Copy-Detection-Pattern (CDP) auf das Ticket aufgebracht. Diese Technologie wurde zur Erkennung von gefälschten Waren entwickelt und wird bei Tabakwaren, Pharmaverpackungen, hochwertigen Ersatzteilen und anderen Markenprodukten eingesetzt [KrVP04]. Durch die Verwendung eines kryptographischen Schlüssels kann das CDP kunden- bzw. produktindividuell erzeugt werden.

Dem CDP-Verfahren und dessen Schutzfunktionalität gegenüber Fälschungen liegt die Erkenntnis zugrunde, dass trotz der nominell hohen Auflösung von Druckern und Scannern sich Abweichungen von der exakten digitalen Darstellung ergeben, die erkennbar sind. Die besondere Herausforderung bei der Umsetzung dieses Lösungsansatzes ist die Klassifizierung der Tickets in Original und Kopie. Hierzu werden komplexe statistische Verfahren eingesetzt, die auch bei der graphischen Mustererkennung und bei der biometrischen Musterklassifizierung erfolgreich eingesetzt werden.

## 4 Integritätsschutz im Detail

### 4.1 Elektronische Signaturen

Die im Barcode gespeicherten Zutrittsberechtigungen müssen gegen Manipulationen geschützt werden. Um hier ein sicheres und standardisiertes Signaturverfahren zu verwenden, werden alternativ das RSA-Verfahren und das Verfahren der elliptischen Kurvenkryptographie gewählt. Als Anbieter für diese kryptographischen Verfahren wurde die CV Cryptovision GmbH in Gelsenkirchen gewählt, da dieses Unternehmen die notwendige Expertise in diesen Signaturverfahren besitzt. Für die Verifikation mittels Handheld ist zudem eine Hardwareimplementierung notwendig.

Da das Verfahren der elliptischen Kurven kürzere Schlüssellängen bei gleicher Sicherheit bietet, ist es hier vorzuziehen, da der Barcode nur eine begrenzte Speicherkapazität je Fläche besitzt. Der Barcode soll möglichst klein sein, da neben den Sicherheitsanforderungen insbesondere Design- und Marketingaspekte eine wichtige Rolle spielen.



Abb. 2: Beispielticket mit Barcode und CDP unterhalb des Lichtbildes

### 4.2 Speicherung der Daten mittels 2D-Barcode

Das 2D-Barcodeverfahren dient dazu, die signierten Zutrittsdaten maschinenlesbar und verifizierbar auf das Ticket aufdrucken zu können. Für das 2D-Barcodeverfahren sollen normierte Standardverfahren eingesetzt werden. Im Bereich der 2D-Barcodeverfahren hat sich in den letzten Jahren der Datamatrix-Code als Industriestandard etabliert. Es handelt sich dabei um einen echten Matrix-Code mit Reed-Solomon-Fehlerkorrektur (aktuell ECC200). Die entsprechende internationale Normierung ist unter der Norm: ISO/IEC 16022 und ANSI/AIM BC 11 verfügbar. Beispiele für diese Codeart sind die elektronische Briefmarke (STAMPIT), der 2D-Pharmacode, Teilekennzeichnungen in Luft- und Raumfahrtindustrie (NASA) und in der Medizintechnik. Ebenso ist diese Codeart auf den Dokumenten der Finanzbehörden bei der Einkommenssteuererklärung zu finden.



Abb. 3: Codebeispiel der Datamatrix als elektronische Briefmarke

## 5 Informationstechnischer Originalitätsschutz

### 5.1 Funktionsweise Copy Detection Pattern (CDP)

Das Copy Detection Pattern stellt ein digitales Sicherheitsmerkmal für Druckerzeugnisse dar, welches alle Anforderungen an hochwertige Sicherheitsmerkmale berücksichtigt, wobei die Wahrnehmbarkeit vom spezifischen Layout abhängt. Das CDP wird in das Ticketlayout integriert und setzt somit vor dem eigentlichen Druckprozess an.

Der Erzeugung eines CDPs liegt folgende Überlegung zu Grunde. Die Aufgabe ist es, ein Muster zu drucken, das möglichst schwer zu reproduzieren ist. Der Ansatz des CDPs geht dabei in zwei Richtungen: Zum einen versucht man, die Kapazität des Papiers als Informationsträger so weit wie möglich auszunutzen. Zum anderen sollte der Versuch, ein solches Muster zu kopieren, zu einem (möglichst zuverlässig) messbaren Informationsverlust führen, der die Unterscheidung von Original und Kopie erlaubt. Aus der Informationstheorie ist bekannt, dass ein (pseudo-)zufälliges Muster den höchsten Informationsgehalt hat. Ein solches Muster lässt sich auch durch Annahmen eines menschlichen Fälschers (wie z. B. über die Regelmäßigkeit einer Form oder Linie, das Vorhandensein von Kanten o. ä. wie etwa bei Guillochen) nicht rekonstruieren. Andererseits enthält ein solches Muster bis zur Nyquist-Frequenz ein von Null verschiedenes Leistungsspektrum, und gerade bei den höchsten Frequenzen ist mit einem Informationsverlust auf der „Übertragungsstrecke“ Drucker-Scanner zu rechnen.

### 5.2 Erstellung eines CDP-geschützten Tickets

Zunächst wird eine kundenspezifische Grafik erzeugt, die ein spezifisches Rauschmuster enthält. Dabei wird bei der Erzeugung ein Schlüssel verwendet, der als Initialwert eines geeigneten Pseudo-Zufallszahlen-Generators verwendet wird. Die Abb. 5 zeigt zwei solcher CDPs, die mit verschiedenen Schlüsseln erzeugt worden sind. Für den menschlichen Betrachter sind sie bei dieser Auflösung nur dann zu unterscheiden, wenn man sich geeignete auffällige Bereiche sucht und diese im anderen CDP vergleicht. Es ist jedenfalls kaum möglich, sich ein CDP zu merken oder ohne direkten Vergleich über die Ähnlichkeit zu entscheiden.

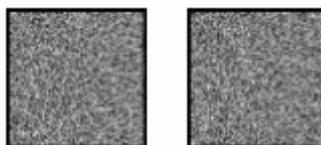


Abb. 4: Zwei CDPs, die mit verschiedenen Schlüsseln erstellt wurden

Dieses Rauschmuster (Grau- oder Farbmuster) kann mit einer anderen Grafik kombiniert werden. Beispielsweise kann das Rauschmuster einem Barcode überlagert werden. Der Barcode (auch 2-dimensionaler Barcode) ist weiterhin lesbar. Ebenso kann das Rauschmuster Grafiken, Bildern oder Veranstalterlogos, die bereits auf dem Ticket enthalten sind, überlagert werden. Die Abb. 6 zeigt Beispiele für solche Überlagerungen von 1-dimensionalen und 2-dimensionalen Barcodedaten und einer Grafik.

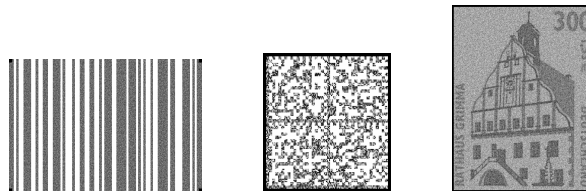


Abb. 5: Überlagerung von Barcodes und einer Grafik mit einem CDP

Durch die Verwendung des Schlüssels kann das CDP kunden- bzw. produktindividuell erzeugt werden. Das entsprechende CDP wird in das Ticketlayout integriert und so automatisch während des herkömmlichen Druckprozesses auf das Ticket aufgebracht. Dabei bestehen nach unserer Erfahrung kaum Beschränkungen im Druckverfahren. Weil das CDP in die digitale Vorlage im pre-press-Bereich eingebracht wird, entstehen nur geringe Integrationskosten und keine zusätzlichen Produktionskosten. In Fällen, in denen das Ticket vorproduziert wird und dies evtl. in nicht zuverlässiger Umgebung bzw. bei gefährdetem Transport vom Drucker zur Personalisierung passiert, kann das CDP am Ort der Personalisierung nachträglich aufgedruckt werden.

Um Fälschungswege zu erkennen, wird zusätzlich in das CDP oder in das Ticketlayout ein digitales Wasserzeichen eingebettet, welches für den menschlichen Betrachter nicht wahrnehmbare Informationen über den Inhaber des Originaltickets enthält. Die Wasserzeicheninformation bietet eine einfache Möglichkeit, Fälschungsquellen zurückzuverfolgen.

Dem CDP-Verfahren und dessen Schutzfunktionalität gegenüber Fälschungen liegt die Erkenntnis zugrunde, dass trotz der nominell hohen Auflösung von Druckern und Scannern sich Abweichungen von der exakten digitalen Darstellung ergeben, die messbar sind. Diese Abweichungen werden insbesondere durch die Mechanik hervorgerufen, die beim Drucken und Scannen – allerdings deutlich stärker beim Drucken – maßgeblich beteiligt sind.

Die Abweichungen zwischen digitalem Original, Ticketdruck und eingelesenem Rasterbild werden durch einen Distanzwert quantifiziert und sind spezifisch für die beteiligten Gerätetypen. Die mehrfache Anwendung dieser Abweichungen bei der Erstellung einer unzulässigen Kopie führt zur Unterscheidbarkeit einer solchen Kopie vom Originalausdruck. Die besondere Herausforderung bei der Umsetzung dieses Ansatzes ist die Klassifizierung der Graphik in Original und Kopie. Hierzu werden komplexe statistische Verfahren eingesetzt, die auch bei der graphischen Mustererkennung erfolgreich eingesetzt werden.

Das Ticket wird zu diesem Zweck mittels CCD-Chip digitalisiert. Das CDP wird lokalisiert und das eingelesene Rasterbild analysiert. Für die Analyse des CDPs wird mit Hilfe eines geeigneten, korrelationsbasierten Abstandsmaßes ein "Distanzwert" zwischen dem digitalen Original und dem vorliegenden, eingelesenen Rasterbild berechnet. Das digitale Original kann dabei entweder als Datei vorliegen oder vor dem Vergleich bei Kenntnis des bei der Erstellung verwendeten Schlüssels erneut berechnet werden. Der Distanzwert ist in der üblichen Weise normalisiert, d.h., ein Wert von 100 entspricht perfekter Übereinstimmung von Original und Kopie (was nur im digitalen Bereich der Fall sein kann). Ein Wert von 0



entsteht beim Vergleich zweier orthogonaler, also mit verschiedenem Schlüssel, erstellter Vorlagen. Unterschreitet dieser Distanzwert einen in der Anwendung festgelegten Schwellwert, handelt es sich um eine unzulässige Kopie, andernfalls um einen Originalausdruck. Dieser Schwellwert sollte in Abhängigkeit von den Eigenschaften des Druck- und Verifikationsprozesses festgelegt werden.

Aus dem hier Gesagten ergibt sich auch, dass sowohl der zur Erstellung eines CDPs verwendete Schlüssel als auch das Original in digitaler Form äquivalent sind und in gleicher Weise vor Kenntnisnahme durch Unbefugte geschützt werden müssen. In dieser Hinsicht ist ein CDP mit einem symmetrischen Kryptographie-Verfahren vergleichbar. Das entsprechende Schlüsselmanagement ist dementsprechend eine wichtige Aufgabe im Rahmen des Entwurfs und der Umsetzung.

Das CDP wurde ursprünglich für den Verpackungsschutz entwickelt und wird in diesem Einsatzgebiet erfolgreich eingesetzt. Für die Evaluierungsergebnisse und Sicherheitsbetrachtungen sei auf das Arbeitspapier [KrVP04] verwiesen, das den Einsatz des CDPs im Verpackungsschutz behandelt. Hierbei wurde die Sicherheit des Verfahrens auch gegenüber professionellen Fälschern nachgewiesen. So wurde in einem Experiment ein CDP im Offset-Verfahren gedruckt. Dann wurde ein Original mit einem Trommelscanner von 10.000 dpi Auflösung digitalisiert und im sog. *computer-to-plate* (CtP)-Verfahren direkt auf der gleichen Offset-Maschine ausgedruckt. Uns wurden dann von Original und Kopie jeweils zehn Exemplare in zufälliger Reihenfolge vorgelegt (Blindversuch). Die Analyse ergab eine 100% Unterscheidung mit einem Abstand der beiden Verteilungen von mehreren Sigma.

## 6 Verifikation der Integrität und Originalität

Es gibt eine Vielzahl mobiler Geräte, die eine Überprüfung solcher Tickets erlauben. Das Gerät muss nur in der Lage sein, das Ticket mit dem Barcode und dem CDP zu digitalisieren, das Ergebnis auszuwerten und eine Rückmeldung zu geben. Als Referenzimplementierung wurde die Funktionalität in das Lesegerät Code Reader MAH 200 der Firma Code Cooperation implementiert.



Abb 6: Code Reader MAH 200

Das Gerät von der Größe eines Mobiltelefons muss nur mit seiner CCD-Leseinheit gegen das Ticket gehalten werden und bestätigt bzw. widerlegt über visuelle und akustische Rückmeldungen die Echtheit des Tickets. Hierzu liest es den Barcode ein und überprüft die Signatur der Zutrittskontrolldaten. Ist die Signatur korrekt und ist der Zutritt laut Datenmaterial erlaubt, so wird der Zutritt gewährt. Das Gerät wird dazu für einen speziellen Übergangsbereich und die aktuelle Spielpaarung vorpersonalisiert und erlaubt dem Kontrolleur eine unmittelbare Entscheidung, ob er den Zutritt erlaubt oder verweigert. Die

Rücknahme von Akkreditierungen ist über Sperrlisten realisiert, die bei der Personalisierung des Gerätes für den Spieltag ebenfalls auf das mobile Verifikationsgerät aufgespielt werden.

Neben der Auswertung des Barcodes wird das mittels CCD-Chip eingelesene CDP analysiert. So kann unmittelbar entschieden werden, ob die Zutrittsdaten nicht nur korrekt sind, sondern ob es sich bei dem Ticket um ein Originalticket handelt.

Zusammengefasst kann somit festgehalten werden, dass alle in Kapitel 2 aufgelisteten Anforderungen an ein wirksames und kostengünstiges Zutrittskontrollverfahren erfüllt werden.

Die Tickets können an jedem Veranstaltungsort dezentral und kostengünstig innerhalb etablierter Druckprozesse erzeugt werden. Es werden Standardverfahren für die Barcodetechnologie und das Signaturverfahren eingesetzt. Das Ticket trägt personenindividuelle Merkmale zur Identifizierung der Personen. Es ist nur ein Ticket pro Person notwendig, welches für alle Veranstaltungen gilt. Die Robustheit wird durch eine Laminierung des Ausdrucks erreicht. Durch die eingesetzten Signaturverfahren in Verbindung mit der Barcodetechnologie wird die Datenintegrität sichergestellt, das CDP sichert die Originalität. Die Verifikation ist kostengünstig, mobil einsetzbar und ermöglicht eine eindeutige und sichere Entscheidung. Es ist keine besondere Expertise der Kontrolleure notwendig. Die Verwendung von Sperrlisten ermöglicht das Zurückziehen der Akkreditierung nach der Ausgabe des Tickets.

## Literatur

- [DieW02] Die Welt. Gefälschte Medikamente in deutschen Apotheken – Rückruf eines Antibiotikums – Auch in den Industrieländern werden Plagiate von Arzneimitteln zu einem immer größeren Problem. Ausgabe vom 01.10.2002.
- [Jare04] Peter Jaret. Fake drugs, real threat. LA Times, 09.02.2004. <http://www.latimes.com/la-he-counterfeit9feb09.htmlstory>
- [KrVP04] R. Krüger, J.C. Vorbrüggen; J. Picard: Digitales Sicherheitsmerkmal für fälschungssichere Verpackungen. In: P. Horster (Hrsg): Elektronische Geschäftsprozesse, 2004, ISBN 3-00-0113137-X
- [SPIE03] DER SPIEGEL. Pharmaindustrie: Der Boom gefälschter Medikamente alarmiert Verbraucher, Produzenten und Politiker. Nr. 30/2003 (21.07.2003), S. 56-58.

## Enterprise Document Security

Vertr. Prof. Dr. Ursula Oesing, Phone: ++49.201-43752-48, Fax: ++49.201-43752-70,  
E-Mail: Ursula.Oesing@FH-Duesseldorf.de, FH Düsseldorf, Josef-Gockeln-Str. 9, 40474 Düsseldorf

Roland Krüger, Phone: ++49.201-43752-70, Fax: ++49.201-43752-70,  
E-Mail: rKrueger@MediaSec.de, MediaSec Technologies GmbH, Berliner Platz 6-8, 45127 Essen

Dr. Ulrich Pinsdorf, Phone: ++49. 6151-155-533, Fax: ++49. 6151-155-499,  
E-Mail: ulrich.pinsdorf@igd.fraunhofer.de, Fraunhofer IGD, Fraunhoferstr. 5, 64283 Darmstadt

### Abstract

Informationen sind für die meisten Unternehmen erfolgskritische Ressourcen. Die mutwillige oder versehentliche Veröffentlichung vertraulicher Informationen ist in der Regel mit einem Schaden für das Unternehmen verbunden. Daher werden in zunehmendem Maß Dokumenten-Management-Systeme (DMS) eingesetzt, die über ein Rechtemanagement einen gewissen Schutz gegen eine ungewollte Verbreitung von Informationen bieten. Das in diesem Artikel beschriebene Softwaresystem bietet durch die Markierung von Dokumenten einen Schutz über die Wirkungsweise eines DMS hinaus und erhöht dadurch die Informationssicherheit des Unternehmens. Dokumente werden beim Auschecken aus dem DMS mit digitalen Wasserzeichen markiert, die geeignete Metainformationen transportieren. So können beispielsweise der Autor und die Vertraulichkeitsstufe auch außerhalb des DMS über diese Metainformationen festgestellt werden. Über Filtermechanismen kann verhindert werden, dass die Dokumente das Unternehmen ungewollt verlassen. Zusätzlich ist es durch die Verwendung von mobilen Agenten im Intranet oder von Suchmaschinen im Internet im Nachhinein möglich, markierte Dokumente wiederzufinden und gemäß einer definierbaren Sicherheitspolitik zu reagieren.

**Keywords:** Security, Privatsphäre, Sicherheitspolitik, digitale Wasserzeichen, Unternehmenssicherheit

## 1 Hintergrund

Wer kennt nicht die Situation, dass man Zugriff auf Dateien erhält, deren Inhalte eigentlich nicht für einen selbst bestimmt sind? Oft stecken hinter solchen Vorfällen Unachtsamkeiten im Umgang mit vertraulichen Daten. Beispiele hierfür sind temporäre Freigaben, die nicht wieder rückgängig gemacht wurden oder die Zwischenspeicherung von Dateien zum Dokumentenaustausch.

Die Vertraulichkeit von Unternehmensdaten wird heutzutage intern im Allgemeinen über eine Rechteverwaltung des Dateisystems sichergestellt, im Bereich der Privatsphäre häufig mit zusätzlicher Verschlüsselungstechnologie. Es liegt daher im Verantwortungsbereich der zugriffsberechtigten Personen, sorgfältig mit Dateien umzugehen und sie in entsprechend geschützte Bereiche abzulegen. Den einzelnen Mitarbeiter überfordert dies aber oft bei seinen täglichen Arbeiten. Zum einen muss der Mitarbeiter die Rechteverwaltung überblicken, wobei sich die Rechtevergabe jederzeit für ihn unbemerkt ändern kann. Zum anderen werden von Anwendungsprogrammen automatisiert Sicherheitskopien bzw. temporäre Dateien erzeugt, deren Speicherorte nicht transparent sind.

Weiterhin ist dem einzelnen Mitarbeiter häufig der Vertraulichkeitsstatus von Informationen nicht bekannt bzw. nicht unmittelbar erkennbar. So werden regelmäßig vertrauliche Dokumente über Internetauftritte veröffentlicht. Das Internetportal der ZDF-Heute-Nachrichtensendung veröffentlichte am 19.07.2005 eine entsprechenden Nachricht<sup>1</sup>, dass nach einer Studie die Suchmaschine Google allein für die Anfragen „confidential“ oder „for

---

<sup>1</sup> <http://www.heute.de/ZDFheute/inhalt/11/0,3672,2340139,00.html>

internal use“ mehr als 19.000 Treffer liefert<sup>2</sup>. Solche Dokumente landen zumeist aus Unachtsamkeit oder Versehen in einem Verzeichnis des Webservers, ohne dass ein Verantwortlicher der Firma dies zur Kenntnis nimmt.

In diesem Artikel werden technische Hilfestellungen beschrieben, die Mitarbeiter bei ihrer Verantwortung für die Vertraulichkeit von Informationen unterstützen und Nachlässigkeiten bei dieser Aufgabe beheben. Die vorgestellte Lösung wird zur Zeit innerhalb des Forschungsprojektes SicAri (<http://www.SicAri.de>) entwickelt und dient dem Schutz der Privatsphäre, der Wahrung der Vertraulichkeit von Unternehmensdaten innerhalb des Intranets und dem Auffinden von „versehentlich“ im Internet veröffentlichten Dokumenten. Der Fokus dieses Textes liegt in der Umsetzung der Markierungsstrategie. Methoden zur Markierung von Dokumenten mittels digitaler Wasserzeichen und dem Auffinden markierter Dokumente im Intranet mittels mobiler Agenten wurden bereits in [OKP04] vorgestellt. Auf die grundlegenden Technologien digitaler Wasserzeichen und mobiler Agentenplattformen wird daher hier nicht erneut eingegangen. Die Literaturliste gibt einen Überblick zu relevanten Publikationen die beide Technologien näher beschreiben. Vielmehr wird zunächst die grundlegende Idee kurz zusammengefasst, das zugrundeliegende Sicherheitsmodell dargelegt und auf die Option eines Integritätsschutzes eingegangen. Danach werden die Markierungs- und Filterstrategien vorgestellt. Im Ausblick wird auf die Nutzung der entwickelten Markierungsalgorithmen zum Urheberrechtsschutz eingegangen.

## 2 Markierung von Dokumenten

Die grundlegende Idee ist, die Zugriffsrechte und den Vertraulichkeitsstatus einer Datei (und den darin enthaltenen Informationen) unmittelbar mit dem Inhalt dieser Datei zu verknüpfen. Bei der Markierung von Dokumenten werden Metainformationen innerhalb eines Dokuments gespeichert. Es ist daher keine zusätzliche Datei erforderlich, in der die Markierung gespeichert ist. Diese Informationen sind transparent in das Dokument integriert, so dass sie für den Benutzer unsichtbar sind und von Anwendungsprogrammen ignoriert werden. Markierungen enthalten als Metadaten Informationen über den Vertraulichkeitsstatus des zugehörigen Dokuments, dienen aber nicht der Zugriffskontrolle auf das Dokument selbst. Die Art der Markierung ist dokumentspezifisch, wobei i.a. beliebig viele Informationen untergebracht werden können. Zur Markierung werden steganographische Verfahren eingesetzt.

### 2.1 Sicherheitsmodell

Zum Schutz der Vertraulichkeit von Dokumenten werden i.a. entsprechende Zugriffsrechte des Dateisystems benutzt. Beim Datenaustausch werden zudem geeignete Verschlüsselungstechnologien verwendet. Ein berechtigter Benutzer kann aber in beiden Situationen auf das Dokument zugreifen und hat unmittelbar Zugang auf die enthaltenen Informationen. Somit kann ein berechtigter Benutzer diese Informationen auch proaktiv veröffentlichen. Hat ein solcher Benutzer die Leserechte auf ein vertrauliches Dokument, so kann er sich dieses Dokument anschauen. Er kann den Inhalt beispielsweise auf ein Diktiergerät sprechen oder mit der im Mobiltelefon integrierten Kamera einen Screenshot erzeugen und das Dokument am privaten PC neu verfassen. Das neue Dokument ist ungeschützt – weder sind entsprechende Zugriffsrechte gesetzt noch ist es verschlüsselt.

---

<sup>2</sup> [http://www.katastrophenschutz.steiermark.at/cms/dokumente/10044203/e0cc1579/mobilkom-austria\\_Pr%E4sentation-Notfallpl%E4ne-2003.ppt](http://www.katastrophenschutz.steiermark.at/cms/dokumente/10044203/e0cc1579/mobilkom-austria_Pr%E4sentation-Notfallpl%E4ne-2003.ppt)

Das hier vorgestellte Softwaresystem dient **nicht** dem Schutz der Vertraulichkeit von Dokumenten, die ein solcher Benutzer **aktiv** als Angreifer veröffentlichen will. Daher ist es auch nicht notwendig, die durch das hier beschriebene System angebrachten Markierungen gegen Manipulationen besonders zu schützen. Manipulationen an den Markierungen führen dazu, dass ggf. die Vorteile des hier vorgestellten Softwaresystems nicht mehr nutzbar sind. Das System hat vielmehr die Aufgabe, das Intranet bzw. Internet nach markierten Dokumenten zu durchsuchen und entsprechend „aufzuräumen“. Findet das System ein markiertes Dokument, entscheidet es nach einer definierbaren Sicherheitspolitik, wie es reagieren soll. Die Reaktion ist sozusagen der Aufräumvorgang, den das System bieten kann. Im Intranet können mobile Agenten von Rechner zu Rechner wandern und falsch abgelegte Dokumente verschlüsseln, löschen oder etwa in ein Sicherungsverzeichnis verschieben. Im Internet bestehen diese Zugriffsmöglichkeiten nicht. Das System hat über die integrierten Metainformationen die Möglichkeit, den Autor zu informieren. Dazu wird als Metainformation beispielsweise die E-Mail-Adresse des Autors in das Dokument integriert. Die Aktionsmöglichkeiten sind allerdings im Allgemeinen auf das lokale Intranet beschränkt, da nur hier seitens der Administration den Agenten die entsprechenden Rechte zugestanden werden. Da sich die mobilen Agenten durch einen Integritätsschutz (Code-Signaturen) auch vertrauenswürdig in anderen Umgebungen bewegen können, kann der Schutz ausgeweitet werden.

## **2.2 Integritätsschutz als zusätzliche Option**

Um die Unversehrtheit und Dokumentenzugehörigkeit einer Markierung zu gewährleisten, können markierte Dokumente, zusätzlich zu der eigentlichen Markierung, digital signiert werden. Eine Änderung am Dokument oder der Markierung kann anschließend durch die Signatur festgestellt werden. Eine Manipulation am Dokument kann aber durch eine digitale Signatur nicht verhindert werden. Es sei hier nochmals auf das zugrundeliegende Sicherheitsmodell verwiesen, dass auch die Markierung von Dokumenten keinen Schutz gegen mutwillige Verbreitung von sicherheitskritischen Daten bieten kann und soll. Die Markierung zusammen mit der mobilen Agentenplattform bietet einen Schutz gegen Unachtsamkeit von Benutzern, nachlässig programmierte oder getestete Anwendungsprogramme und Fehler bei der Administration der Rechteverwaltung des Dateisystems.

## **2.3 Wirkung und Vorteile der Vorgehensweise**

Mit Hilfe dieser Markierungen ist es möglich, die Zugriffsrechte und den Vertraulichkeitsstatus einer Datei explizit und unabhängig vom Speicherort zu erkennen. Dies gilt auch für alle Kopien der Datei. Bei einer Rechteverwaltung über ein Dateisystem kann dies bisher nur implizit über den Speicherort erfolgen.

Durch entsprechende Suchalgorithmen können über die Dokumentenmarkierung Widersprüche zwischen den intendierten und den aktuellen Zugriffsrechten erkannt werden, welche anschließend gemäß einer Sicherheitspolitik aufgelöst werden können.

Ein geeigneter Ansatz zum Schutz von Unternehmensdaten ist die Erweiterung von Dokumenten-Management-Systemen (DMS) um Methoden zur Dokumentenmarkierung. So kann die Vertraulichkeit von Dokumenten über den Wirkungsbereich eines DMS hinaus geschützt werden. In der Markierung können fast beliebige Informationen über den aktuellen Benutzer, den im DMS festgelegten Sicherheitslevel oder auch Kontaktdaten des Autors enthalten sein. Anhand solcher Markierungen können vertrauliche Dokumente auch außerhalb eines DMS aufgefunden werden. Gemäß einer frei definierbaren Sicherheitspolitik

wird dann entschieden, ob die Fundstelle hinreichend vertrauenswürdig ist. Ist dies nicht der Fall, wird entsprechend der Sicherheitspolitik und den technischen Möglichkeiten verfahren. Die Mobile Agentenplattform bietet beispielsweise die Möglichkeit, das vertrauliche oder private Dokument zu verschlüsseln, es in ein als vertrauenswürdiges gekennzeichnetes Verzeichnis zu verschieben oder eine Benachrichtigung an den Autor zu versenden.

Dokumentenmarkierungen helfen somit, ein lokales Intranet bzgl. das Rechtemanagement konsistent zu halten und die versehentliche Verbreitung von sicherheitskritischen Dokumenten einzudämmen. Ebenso können vertrauliche Dokumente, die versehentlich im Internet veröffentlicht wurden, über Suchmaschinen aufgefunden werden. Eine weitere Verbreitung vertraulicher Informationen kann so eingedämmt werden – hoffentlich bevor eine breite Öffentlichkeit hiervon Kenntnis erlangt.

### **3 Markierungs- und Suchstrategien**

Im Folgenden werden anhand möglicher Einsatzszenarien verschiedene Strategien zur Einbettung von Dokumentenmarkierungen und dem Auffinden markierter Dokumente vorgestellt. Es werden unterschiedliche Sicherheitspolitiken exemplarisch dargelegt. Diese Politiken definieren die Reaktionen auf erkannte Sicherheitsvorfälle, d.h. auf Widersprüche zwischen der Vertraulichkeitseinstufung des aktuellen Speicherorts einer Datei und deren Vertraulichkeitsmarkierung.

#### **3.1 Privatsphäre**

Private Dokumente werden im Allgemeinen in besonders geschützten Bereichen abgelegt. Als potentielles Risiko sind hier Fehler in der Rechteverwaltung des Dateisystems und lokale Sicherungskopien zu nennen. Zudem dienen Dokumente häufig der Kommunikation. Auf den Schutz persönlicher Dokumente beim Kommunikationspartner hat man meistens keinen Einfluss, ebenso auf die Erstellung von Kopien. Daher wird in persönliche Dokumente der Vermerk *persönlich* eingetragen. Die Dokumentenmarkierung enthält neben diesem Vermerk auch die Kontaktdaten und das Verschlüsselungszertifikat des Erstellers, so dass beim Auffinden eines persönlichen Dokuments der Ersteller benachrichtigt werden kann und das Dokument ggf. verschlüsselt wird. Die Suchstrategie entspricht der Strategie zum Auffinden vertraulicher Unternehmensdaten, so dass die technischen Umsetzungen im folgenden Abschnitt vorgestellt werden.

Es ist zu beachten, dass Datenschutzaspekte hier eine wichtige Rolle spielen und ggf. Vereinbarungen mit den Betroffenen geschlossen werden müssen, insbesondere wenn es um die Kontrolle der privaten Nutzung von Unternehmensressourcen handelt.

#### **3.2 Vertrauliche Unternehmensdaten**

Bei vertraulichen Unternehmensdaten kann die Markierung manuell ausgelöst werden. Dies hat den Vorteil, dass die Markierungsalgorithmen bei Bedarf unmittelbar eingesetzt werden können. Allerdings obliegt es der Verantwortung des Mitarbeiters, die korrekten Markierungsoptionen zu setzen, insbesondere überhaupt eine Markierung anzustoßen. Daher ist eine solche Markierungsstrategie nur für besonders sensible Dokumente bei einer überschaubaren Anzahl einsetzbar. In Unternehmen mit größerem Dokumentenaufkommen werden zur Verwaltung der Dokumente sinnvollerweise Dokumenten Management Systeme (DMS) eingesetzt, so dass hier die Markierungen durch eine Erweiterung automatisiert im Hintergrund erfolgen können. Als Referenzimplementierung wurde dieser Ansatz in das Xinco™ DMS integriert (<http://www.xinco.org>). Vorteile dieses DMS sind die freie

Verfügbarkeit und die Plattform-Unabhängigkeit durch die Implementierung in der Programmiersprache Java. Die Markierungsalgorithmen sind dokumententypabhängig. Daher werden bisher Datenformate unterstützt, die i.a. unternehmenskritische Daten enthalten. Dies sind insbesondere die Microsoft Office Datenformate inkl. dem Rich Text Datenformat, Star- bzw. OpenOffice Datenformate, das Adobe Portable-Document-Format (PDF), alle gängigen Bilddatenformate und in XML beschriebene Datenformate.

Um markierte Dokumente im Intranet bzw. Internet aufzufinden, können die Suchstrategien von Internetsuchmaschinen verwendet werden bzw. solche Suchmaschinen um die entsprechenden Leseroutinen erweitert werden [EbPi05]. Im lokalen Intranet bietet sich die Technologie der mobilen Agenten an. Hierzu wurde die mobile Agentenplattform SeMoA™ (<http://www.SeMoA.org>) um eine Suchstrategie erweitert. Über entsprechende Sicherheitspolitiken führen die mobilen Agenten auch die definierten Reaktionen aus, die von einer Benachrichtigung über Verschlüsselungen bis zu einem Löschen der Datei gehen.

Die bisher vorgestellten Technologien (Markierungsalgorithmen, Agentenplattform, DMS) sind durch ihre Java-Implementierung von der eingesetzten Plattform unabhängig. Die in den beiden folgenden Kapiteln vorgestellten Einsatzbeispiele müssen individuell an die verwendete Serverplattform bzw. das eingesetzte Betriebssystem angepasst werden. Es ist zu beachten, dass es sich hier im eigentlichen Sinn nicht um Markierungs- und Suchalgorithmen handelt, sondern um Filterfunktionalitäten, die in bestehende Systeme integriert werden müssen. Daher sind diese Integrationslösungen nur in Abhängigkeit von den eingesetzten Systemen und Plattformen einsetzbar.

### **3.3 Server-Einsatz als Security-Gateway**

Um die versehentliche Verbreitung von sicherheitskritischen Dokumenten zu erschweren, kann die Markierungsplattform auch in Form eines Filter-Plug-Ins in verschiedene Server-Dienste eingebunden werden. Auf einem SMTP-Server können so ausgehende E-Mails auf Dokumente untersucht werden, die nicht für Personen außerhalb eines bestimmten Personenkreises gedacht sind. Wird ein solches Dokument erkannt, so wird es, je nach Sicherheitsstufe, von der Auslieferung zurückgestellt und der Sender per E-Mail benachrichtigt oder der Versand komplett abgelehnt. Soll das Dokument trotzdem verschickt werden, so muss der Sender auf die E-Mail des Systems reagieren, damit sicher gestellt ist, dass das Dokument nicht versehentlich verschickt werden sollte. Diese Lösung ist, um die Serverlast klein zu halten, auch auf lokalen Rechnern möglich. Hier wird die Markierungsplattform als Proxy-Server eingesetzt, der alle ausgehenden Verbindungen auf Port 25 (SMTP) überwacht und die zu sendenden Daten zuerst abfängt und auswertet.

Der Einsatz auf einem Web- oder FTP-Server verläuft ähnlich zu dem auf einem SMTP-Server. Anfragen an bestimmte Dokumenttypen werden hier zunächst durch das Plug-In abgefangen. Wurde ein Dokument angefordert, das aufgrund seiner Sicherheitsstufe nicht auf einem Verzeichnis des Servers liegen sollte, so wird es in ein geschütztes Verzeichnis verschoben. Ein Benutzer erhält dann als Antwort auf die Anfrage eine Fehlermeldung für eine nicht gefundene Ressource. Zusätzlich wird der Administrator oder eine andere zuständige Instanz über den Zwischenfall informiert.

### **3.4 Einsatz im lokalen Dateisystem**

Ähnlich wie ein Virenschanner kann auch die Markierungsplattform in ein Betriebssystem eingegliedert werden. Wenn ein neues Dokument erstellt wird, so kann es automatisch mit einer Markierung versehen oder bestehende Dokumente auf das Vorhandensein von Meta-

Informationen untersucht werden. Dazu muss jedes Verzeichnis auf einem Rechner ebenfalls eine Sicherheitsstufe besitzen, damit geprüft werden kann, ob ein Dokument an dieser Stelle abgelegt werden darf oder nicht. Dies wird in einer Datenbank der Markierungsplattform gespeichert. Möglich ist in diesem Fall auch, dass die Speicherung eines sicherheitskritischen Dokuments an bestimmten Stellen im Dateisystem komplett verhindert wird. Beispielsweise wäre es nicht möglich, den Arbeitsvertrag eines neuen Mitarbeiters in einem Verzeichnis zu speichern, das für den internen Datenaustausch vorgesehen ist.

### 3.5 Implementierung eines Anwendungsbeispiels

Um das Zusammenwirken der Einzelkomponenten (Markierungsalgorithmen, Xingo-DMS als Markierungsplattform, SeMoA als Such- und Reaktionsplattform) zu demonstrieren, wurden Anwendungsumgebungen nachempfunden. Exemplarisch wird in diesem Artikel das Szenarium „Berufungskommission“ vorgestellt, da bei der Berufung eines Professors viele sensible Dokumente erzeugt und ausgetauscht werden. Zudem besitzen die Mitglieder einer Berufungskommission unterschiedliche Berechtigungen und Vertrauensstellungen. So haben Hochschulangehörige mit unterschiedlichen Befugnissen unterschiedliche Zugriffsberechtigungen auf die Dokumente, die innerhalb eines Berufungsverfahrens erzeugt und verteilt werden. Insbesondere die erstellten Gutachten sind hier in bezug auf ihre Vertraulichkeit besonders sensibel. Im folgenden wird ein einfaches Beispiel vorgestellt, das aber bereits viele Funktionalitäten der vorgeschlagenen Lösung nutzt.

Ein Professor arbeitet bevorzugt auf der Desktopoberfläche seines Rechners und erstellt ein Gutachten über einen Bewerber. Für solche Gutachten steht ihm ein eigener Bereich über das DMS zur Verfügung, in das er Gutachtendateien hochlädt. Die Nutzerrechte dieses Verzeichnisses sind restriktiv. Außerdem werden alle dort gespeicherten Dateien vollautomatisch so verschlüsselt, dass sie nur mit einem auf einer Smartcard befindlichen Schlüssel wieder lesbar gemacht werden können. Die einzige Kopie dieses Schlüssels ist auf einer Diskette in seinem Tresor abgelegt. Auf dem DMS speichert er den ersten Entwurf eines neuen Gutachtens. Solange er die aktuelle Version noch auf der Desktopoberfläche bearbeitet, wird nichts verschlüsselt. Das DMS markiert aber bei der Erstellung des Dokumentes die Datei als vertrauliches Gutachten. Der Professor muss wegen eines Termins das Büro verlassen und vergisst in seiner Eile das Gutachten im DMS abzulegen, sondern loggt sich nur aus.

Die SeMoA-Agentenplattform aktiviert nun zu einer festgelegten Zeit bzw. in Abhängigkeit von der aktuellen Nutzlast einen Suchagenten, der die Verzeichnisse aller Arbeitsplatzrechner untersucht. Im Desktop-Verzeichnis wird das ungesicherte Gutachten gefunden und es werden der Reihe nach folgende Aktionen durchgeführt, die so in der Sicherheitspolitik definiert wurden:

- die Datei wird verschlüsselt,
- die Datei wird in ein gesichertes Verzeichnis verschoben,
- der Professor wird informiert.

Damit niemand außer dem Professor die Datei wieder entschlüsseln kann, wird das Verschlüsselungszertifikat des Professors für die Verschlüsselung des Dokuments verwendet. Zunächst wird über einen für kryptographische Anwendungen geeigneten Zufallszahlengenerator ein symmetrischer Verschlüsselungsalgorithmus wie DESede oder AES initialisiert, mit dem das Dokument verschlüsselt wird. Anschließend werden die hierfür benötigten Parameter mit dem Public Key aus dem Verschlüsselungszertifikat des Professors verschlüsselt und dem Dokument hinzugefügt (Hybridverschlüsselung).



Zusätzlich wird das Dokument in ein als „sicher“ gekennzeichnetes Verzeichnis verschoben und der Professor per E-Mail über den Zwischenfall informiert. Ein sicheres Verzeichnis kann ein Speicherort innerhalb des DMS oder eine beliebige andere Stelle auf dem betreffenden Rechner oder im Netzwerk sein. Selbstverständlich sind alle Aktionen der Agentenplattform ebenfalls kryptographisch abgesichert, sei es die Authentisierung der Agenten gegenüber dem Rechner des Professors, der Integritätsschutz des mobilen Codes oder beispielsweise die verschlüsselte Kommunikation zu anderen Rechnern beim Datentransfer.

## 4 Ausblick

Neben den beschriebenen Einsatzgebieten ist die Lösung auch als Filter-Plug-In für PrintServer zur Kontrolle von Ausdrucken einsetzbar. Desweiteren werden zur Zeit Dokumentenmarkierungen für Geodaten (GIS), technische Zeichnungsformate (CAD) und multimediale Daten auf Basis einer Wasserzeichentechnologie entwickelt. Weitere Dokumentenformate sind möglich und werden bei Bedarf unterstützt.

Neben dem Schutz der Vertraulichkeit durch die Nachverfolgung von Distributionswegen können die Dokumentenmarkierungen auch im Urheberrechtsschutz eingesetzt werden. Hierzu sind aber Dokumentenmarkierungen notwendig, die derart mit dem Dokument verwoben sind, dass sie nur schlecht entfernbar sind oder zu einer Zerstörung bzw. wesentlichen Qualitätsverschlechterung führen. Insbesondere neue, intelligente Wasserzeichenverfahren werden zur Zeit entwickelt. Bei Audio- und Videoformaten sind die zu lösenden Aufgaben allerdings besonders groß, da das Datenmaterial oft durch die Komprimierungsverfahren sehr wenige Freiheiten für die Einbettung von Metainformationen zulässt. Zudem entwickeln sich diese Multimediaformate in einem rasanten Tempo, so dass die Markierungsalgorithmen häufig angepasst werden müssen. Hier zeichnen sich aber vielversprechende Lösungen ab, die auch mit den neusten Formaten arbeiten und den Urheberrechtsschutz aktiv unterstützen werden.

Innerhalb des Forschungsprojektes SicAri wird zur Zeit untersucht, ob Wasserzeichenalgorithmen nicht nur für Dokumentenmarkierungen eingesetzt werden können, sondern auch den Integritätsschutz von Dokumenten erweitern können. Insbesondere werden die Algorithmen zum Urheberrechtsschutz untersucht und weiterentwickelt. Die grundlegende Idee hierbei ist, das gesamte Dokument redundant derart zu kennzeichnen, dass eine partiell fehlende Markierung Rückschlüsse auf eine Manipulation erlaubt. Vorteil solcher Integritätsschutzverfahren ist, dass sie beispielsweise auch nach Formatsänderungen oder für Teilausschnitte gültig bleiben.

## Literatur

- [ArSW03] M. Arnold, M. Schmucker, S. Wolthusen: Techniques and Applications of Digital Watermarking and Content Protection, Artech House (2003), ISBN 1-58053-111-3.
- [CKLS97] I.J. Cox, J. Kilian, T. Leighton, T. Shamoan: Secure Spread-Spectrum Watermarking for Multimedia, IEEE Transactions on Image Processing, 1997.
- [CKM+01] R. S. Cost, S. Kallurkar, H. Majithia, C. Nicholas, Y. Shi: Integrating Distributed Information Sources with CARROT II. In: Proceedings of the 6th International Workshop on Cooperative Information Agents VI, 2002, ISBN 3-540-44173-5.
- [EbPi05] P. Ebinger, U. Pinsdorf: Automatisierte Recherche von Markenpiraterie im Internet. In: P. Horster (Hrsg.): D-A-CH Security 2005: Bestandsaufnahme, Konzepte, Anwendungen, Perspektiven, 2005, ISBN 3-0001-5548-1.
- [LaOs98] D. B. Lange, M. Oshima: Programming and Deploying Java Mobile Agents with Aglets. Peachpit Press (1998), ISBN 0201325829.
- [LaOs99] D. B. Lange, M. Oshima: Seven Good Reasons for Mobile Agents. In: Communications of the ACM, Vol. 42, Nr. 3, 1999.
- [OKP04] U. Oesing, R. Krüger, U. Pinsdorf: Distributionskontrolle von Dokumenten mittels Wasserzeichen und mobilen Agenten. In: P. Horster (Hrsg): Elektronische Geschäftsprozesse, 2004, ISBN 3-00-0113137-X
- [PiBu03] U. Pinsdorf, Ch. Busch. Neue Potenziale für mobile Dienstleistungen – Einsatz von mobilen Agenten im mobilen Arbeitsumfeld. In: M. Weiss, Ch. Busch, W.Schröter (Hg.): Multimedia Arbeitsplatz der Zukunft. Assistenz und Delegation mit mobilen Softwareagenten. Talheimer Verlag (2003). ISBN 3-89376-105-5.
- [PPHG02] U. Pinsdorf, J. Peters, M. Hoffmann, P. Gupta: Context-aware Services based on Secure Mobile Agents. In: Proceedings of 10th International SoftCOM Conference, Oktober 2002. IEEE Communication Society. ISBN 953-6114-52-6.
- [RoJa01] V. Roth, M. Jalali: Concepts and Architecture of a Security-centric Mobile Agent Server. In: Proc. Fifth International Symposium on Autonomous Decentralized Systems (ISADS 2001), Dallas, Texas, U.S.A., March 2001. IEEE Computer Society Press.
- [Roth01] V. Roth: Sichere verteilte Indexierung und Suche von digitalen Bildern. Dissertation, Technische Universität Darmstadt (2001).
- [Uspa97] U.S. Patent and Trademark Office: System and Method for Distributed Computation Based upon the Movement, Execution and Interaction of Processes in a Network. US patent no. 5603031.

# **eGovernment für Hochschulen und Studierende - Herausforderungen und Möglichkeiten auch im Bologna-Prozess**

Prof. Dr. Hermann Strack, Phone: ++49.3943-659-307, Fax: ++49.3943-659-399, E-Mail: hstrack@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

Christoph Karich, Phone: ++49.3943-659-370, Hochschule Harz, ckarich@hs-harz.de

Peter Kußmann, Phone: ++49.3943-659-369, Hochschule Harz, pkussmann@hs-harz.de

## **Abstract**

Im Folgenden werden durch Anwendung und Ausbau von eGovernment-Standards wie OSCI, zur datentechnischen Begleitung der EU-weiten Mobilität von Studierenden in Bachelor-/Mastersystemen im Bologna-Prozess, grundlegende Konzepte für die strukturierte und gesicherte Übermittlung und Speicherung von Studierenden-, Studiengangs-, Prüfungs- und Zeugnisdaten erarbeitet (bis hin zu digitalen Hochschul- und Schulzeugnissen), um eine effiziente und gesicherte EU-weite Austauschbarkeit solcher Daten im Sinne von eGovernment für Hochschulen zu erreichen.

**Keywords:** eGovernment, Bologna-Prozeß, Signatur, OSCI, Standards.

## **1 Verwaltung und Datenverarbeitung an Hochschulen**

Das Datenmanagement an Hochschulen lässt sich nach verschiedenen Kategorien und Verantwortungsbereichen unterteilen, einige Kategorien orientiert nach den verantwortlichen Akteuren und deren Aufgaben sowohl „legislativer“ als auch „exekutiver“ Art und Adressaten (im Rahmen externer z.B. gesetzlicher Vorgaben):

- Rechtsverbindliche Außenvertretungen der Hochschule (Rektor, Kanzler)
- Eigenverwaltung von Dozenten, Forschern und Studierenden; internes Management durch interne Organisationseinheiten bzw. Stellen (z.B. Rektorat bzw. Rektor, Gremien, Fachbereiche bzw. Dekan,...)
- Verwaltungsstellen für Immatrikulations-, Studierenden- und Prüfungsdaten; Verwaltungs-/Controlling-Stellen für Budgets (Lehre, Forschung, Stellen,...); sonstige Verwaltungsstellen differenzierbar nach internen bzw. externen Verantwortlichkeiten, Bezügen oder Adressaten.

Vielfach sind Hochschulverwaltungen bezüglich rechtsverbindlicher Interaktionen mit der Außenwelt noch fest in der Papierwelt verankert, die Elektronisierung von Verwaltungsverfahren findet, wenn überhaupt, oft vorrangig im internen Verantwortungs- und Gestaltungsbereich der Hochschulen statt. Durch die Umgestaltung der Hochschul- und Studienstrukturen in Europa durch den Bologna-Prozess entstehen neue Herausforderungen bezüglich des (potentiell EU-weiten) Austausches und zugeordneter Verwaltung von Studierenden- und Prüfungsdaten.

## **2 Bologna-Prozess und Verwaltungsaspekte**

Zur Schaffung eines europäischen Hochschulraumes im Bologna-Prozess haben die beteiligten Länder gemeinsame Ziele zur Restrukturierung von Hochschulen und Studiengängen und –abschlüssen gesetzt [9,10], insbesondere die Einführung international

kompatibler gestufter Studiengänge/Abschlüsse (Bachelor/Master), die Modularisierung von Studiengängen verbunden mit einer verbesserten internationalen Anerkennung von Prüfungs- und Studienleistungen unter Berücksichtigung der modul-spezifischen Belastung der Studierenden durch das ECTS-Leistungspunktesystem und damit verbunden die verbesserte Unterstützung der internationalen Mobilität von Studierenden und Lehrenden. Ferner wird die internationale Transparenz und Kompatibilität von Studienabschlüssen durch die Ausstellung eines Diploma Supplements zusätzlich zum nationalen Hochschulzeugnis gefördert. Unter Verwaltungs- und Datenmanagement-Aspekten sind hier insbesondere die Verfahren zum Austausch von bescheinigten Prüfungs- und Studienleistungen sowie Hochschulabschlüssen von Interesse. Erinnerung sei hier zunächst noch einmal an wichtige Teilziele und Instrumente des Bologna-Prozesses:

- Modularisierung der Studiengänge mit Einführung eines ECTS-kompatiblen Leistungspunktesystems, das die Studienbelastung des Studierenden durch den jeweiligen Modul quantitativ angibt.
- ECTS: durch die Einführung des European Credit Transfer System (ECTS) kann der Arbeitsumfang, den der Studierende für das Erbringen der relevanten Leistung benötigt hat, deutlich gemacht werden. Zusätzlich: Die ECTS-Note macht unabhängig vom nationalen Notensystem deutlich, wie die Leistung des Studierenden innerhalb seines Jahrgangs einzuordnen ist.
- Anerkennung von Prüfungsleistungen und Studierendenmobilität: Studierenden und Absolventen wird somit eine verbesserte internationale Anerkennung ihrer Studien- und Prüfungsleistungen ermöglicht. Dies fördert die gewünschte Mobilität der Studierenden während und nach der Ausbildung. Im Rahmen der Bescheinigung und Anerkennung externer Studienleistungen wird bisher mit Papierformularen gearbeitet (wie dem „Transscript of Records“, kurz „TOR“), die von verantwortlichen Stellen handschriftlich unterschrieben und gestempelt werden, siehe Abbildung 1: Formulare für den (nationalen) Prüfungsdatenaustausch (als TOR) in Sachsen-Anhalt [13]. Probleme bereiten hier die Medienbrüche bei der Übermittlung der Daten zwischen Hochschulen (Ineffizienz) und insbesondere bei supra-nationalen Austauschen die Datensicherheit, insbesondere die Authentizität und Integrität der ausgefüllten Formulare. So existieren nicht einmal Stempelregister der beteiligten Hochschulen im Bologna-Prozess der EU zur Unterstützung der Echtheitsprüfung von solchen TOR-Dokumenten (nach Auskunft der zuständigen Verwaltungsstelle der Hochschule Harz).

NAME DER HEIMATHOCHSCHULE: .....					
Fakultät/Fachbereich: .....					
ECTS Fachbereichsordinator/in: .....					
Tel.: ..... Fax: ..... E-mail: .....					
NAME DES STUDENTEN/DER STUDENTIN: ..... Vorname: .....					
Geburtsdatum und Geburtsort: ..... (Geschlecht) : .....					
Tag der Immatrikulation: ..... Immatrikulationsnummer: .....					
NAME DER GASTHOCHSCHULE: .....					
Fakultät/Fachbereich: .....					
ECTS Fachbereichsordinator/in: .....					
Tel.: ..... Fax: ..... E-mail: .....					
Kurscode (1)	Titel des Kurses	Kursdauer (2)	Note (3)	ECTS-Note (4)	ECTS Anrechnungspunkte (5)
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....
	Gegebenenfalls auf getrenntem Blatt fortführen				Gesamt: .....

(1) (2) (3) (4) (5) Erläuterungen befinden sich auf der Rückseite

Erworbene Diplome/Abschlüsse: .....

Datum

Unterschrift des Dekans/Ltd. Verwaltungsbeamten

Hochschulstempel

Hinweis : Dieses Dokument ist nur mit Unterschrift des Dekans/Ltd. Verwaltungsbeamten und offiziellem Hochschulstempel gültig.

Abb. 1 ECTS-Formular, vgl. [13]: Bescheinigung der Gasthochschule für den Studierenden, einzureichen zur Anerkennung bei der Heimathochschule (als TOR)

### 3 eGovernment – Technik, Sicherheit und Recht

Zur system-, raum- und zeitübergreifenden Sicherung der Authentizität und Integrität von Daten wurden bereits früh kryptographische Sicherheitsfunktionen, namentlich die digitale Signatur entwickelt. In Deutschland wurde eine entsprechende Public-Key-Infrastruktur (PKI) nach Signaturgesetz und Signaturverordnung (SigG/SigV) in 1997 gesetzlich geregelt – Deutschland war damit Vorreiter in Europa. Die Regelungen für die PKI nach SigG/SigV waren dabei von einem hohen Sicherheits- und Qualitätsniveau geprägt. Durch eine Rahmenrichtlinie der EU zu elektronischen Signaturen wurden verschiedene Signaturebenen eingeführt, die inzwischen umgesetzt wurden. Dabei hat die in Deutschland (kurz) sogenannte „qualifizierte Signatur“, genauer „fortgeschrittene elektronische Signaturen, die auf einem qualifizierten Zertifikat beruhen und die von einer sicheren Signaturerstellungseinheit erstellt werden“, neben der hochwertigen sicherheitstechnischen Ausstattung die besondere rechtliche Anerkennung, dass sie „.... die rechtlichen Anforderungen an eine Unterschrift in bezug auf in elektronischer Form vorliegende Daten in gleicher Weise erfüllen wie handschriftliche Unterschriften in bezug auf Daten, die auf Papier vorliegen, und in Gerichtsverfahren als Beweismittel zugelassen sind.“

Zusätzlich wurde in Deutschland für die gesicherte und rechtsverbindliche Kommunikation verteilter Anwendungen im Bereich des eGovernment bzw. eCommerce im 1998 durch das BMWi initiierten und geförderten Städtewettbewerb MEDIA@Komm [5,6] das „Online Services Computer Interface“ (OSCI) [1] entwickelt, welches eine Protokoll- und Datenschemata-Sammlung mit integrierten Sicherheitsfunktionen beschreibt.

OSCI verfolgt dabei mehrere Sicherheitsziele gleichzeitig. Dazu zählen die Vertraulichkeit der Inhaltsdaten, welche über eine Ende-zu-Ende-Verschlüsselung sichergestellt wird, die Integrität und Authentizität, die mit Hilfe verschiedener Arten elektronischer Signaturen umgesetzt werden und die verbindliche Zustellung von Nachrichten, d.h. dass Nachrichtentransporte gesichert quittiert und protokolliert werden müssen. Eine direkte Kommunikation zwischen den Kommunikationspartnern (Ende-zu-Ende) findet bei OSCI nicht statt, sondern erfolgt immer zwingend über einen so genannten OSCI-Intermediär. Dieser erbringt neben der reinen Kommunikationsvermittlung zwischen den Beteiligten so genannte Mehrwertdienste, insbesondere Zertifikatsprüfungen, Verschlüsselungen und signierte Zeitstempelungen. Ferner führt der Intermediär insbesondere für Zustellnachweise einen signierten (so genannten) Laufzettel zum Nachweis über jede OSCI-Nachrichtebearbeitung, in dem spezifische Daten, wie Nachrichteneingang bzw. -ausgang, Zertifikatsprüfungen, etc. gespeichert werden.

Um das Sicherheitsziel der Vertraulichkeit durchgängig abbilden zu können, unterscheidet der OSCI-Standard zwischen Nutz- und Inhaltsdaten. Die Nutzdaten (auch OSCI-Teil-A oder OSCI-Transport) beschreiben dabei lediglich die äusseren Transportprotokollcontainer, welche die darin enthaltenen digital signierten und verschlüsselten Inhaltsdaten-Container eines Fachverfahrens XML-repräsentiert und zusätzlich transportverschlüsselt zum Intermediär zur dortigen Entschlüsselung übertragen. Die ebenfalls XML-repräsentierten Inhaltsdaten-Container (nach fachverfahrensspezifischen XML-Schemata, OSCI-Teil-B/OSCI-XÖV-Standards<sup>1</sup>) dagegen sind Ende-zu-Ende zwischen den Kommunikationspartnern verschlüsselt. Dieses Prinzip wird häufig auch als das Prinzip des „Doppelten Umschlags“ bezeichnet, wie nachfolgende Abbildung verdeutlichen soll.

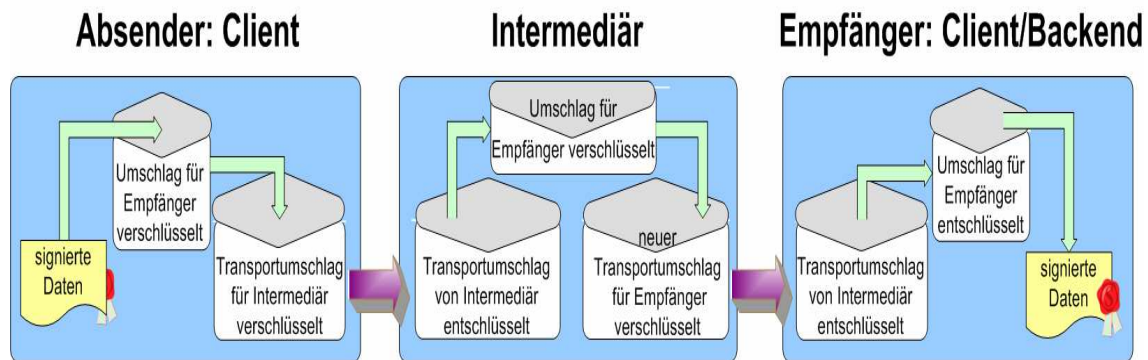


Abb. 2: OSCI-Prinzip „doppelter Umschlag“

Eine OSCI-Kommunikation wird über den SOAP-Internetprotokollstandard abgebildet, eine OSCI-Nachricht (bestehend aus Inhalts- und Nutzdaten) entspricht dabei einem XML-konformen Dokument, welches die Erbringung von Signaturen und Verschlüsselungen über die Standards XML-Signature und XML-Encryption ermöglicht. Darüber hinaus stellt die Verwendung der XML-Technologie die Grundlage dar, die Inhaltsdaten als XML-Schema zu definieren, und somit die Interoperabilität verschiedener Verfahren und Systeme, die auf diese Daten zugreifen, strukturell sicherzustellen. Über Ersetzungen der XML-Inhaltsdatenschemata für die OSCI-Inhaltsdaten-Container, können verschiedene Fachverfahren in OSCI-basiertes eGovernment integriert werden.

<sup>1</sup> XÖV-Standards: XML-basierte Fachverfahrens-Standards für die öffentliche Verwaltung

Als namhafte Vertreter solcher spezifizierten OSCI-XÖV- bzw. Inhaltsdaten-Standards sind XJustiz zur Abbildung des elektronischen Rechtsverkehrs und XMeld [4] zur Abbildung des elektronischen Meldewesens hervorzuheben (bundesweit umzusetzen bis 1.1.2007), die von der OSCI-Leitstelle [1] (wie OSCI-Transport) und entsprechenden fachlichen Projektgruppen im Auftrag des „Kooperationsausschuss Automatisierte Datenverarbeitung Bund/Länder/Kommunaler Bereich“ (KoopA ADV) entwickelt und gepflegt werden. Der Aufbau der bundesweiten OSCI-Infrastrukturen wird durch die jeweiligen eGovernment-Aktivitäten der einzelnen Bundesländer vorangetrieben, welche analog zur eGovernment-Initiative des Bundes „BundOnline 2005“ eGovernment-Infrastrukturen und –Verfahren in eigener Hoheit aufbauen, betreiben und ggf. zur Nachnutzung dem kommunalen Bereich und anderen Stellen anbieten, die der Landesverwaltung nachgeordnet sind. Das Bundesland Sachsen-Anhalt hat dafür den so genannten „eGovernment-Aktionsplan für die Landesverwaltung Sachsen-Anhalt“ [0] veröffentlicht, welcher vorsieht, die notwendigen Basiskomponenten (d.h. Infrastrukturkomponenten), wie z.B. Digitale Signatur sowie Virtuelle Poststelle mit OSCI-Intermediär, Dienstleistungsportal, Content Management System, Formularserver, Zahlungsverkehrsplattform und Geodatenserver zentral bereitzustellen. Hierzu führt die Hochschule Harz im Auftrag des Ministeriums des Innern Sachsen-Anhalt (MI LSA), dort der Landesleitstelle IT/eGovernment (LIT) eine Begleitforschung zur Umsetzung des eGovernment-Aktionsplans Sachsen-Anhalt (BeGovSAH) durch [3] und sprach eine Architekturempfehlung zum Aufbau einer OSCI-konformen Arbeitsumgebung (OSCI-Intermediär) im Lande Sachsen-Anhalt aus, die zum landeszentralen Aufbau der OSCI-Infrastrukturkomponente Intermediär führte, die zur Nachnutzung durch den kommunalen Bereich bereitsteht (insbesondere für die Meldebehörden). Darüber hinaus begleitete das Projekt BeGovSAH den organisatorischen und technischen Aufbau der landeszentralen Public-Key-Infrastruktur (PKI), welche es ermöglicht, den Verwaltungsstellen fortgeschrittene und qualifizierte Zertifikate nach Signaturgesetz auszustellen, die für Verschlüsselungen und rechtskonforme Signaturen benötigt werden.

Von diesen Landesprojekten und den genannten Basiskomponenten sind hohe Synergieeffekte bei der Integration von eGovernment-Prozessen an Hochschulen zu erwarten, da auf die OSCI-Infrastruktur aufgesetzt werden kann, und somit entsprechende Kosten für separate Entwicklungen, Integration und Wartung entfallen und von Erfahrungen bei der Implementation von eGovernment-Prozesse anderer Stellen Gebrauch gemacht werden kann.

Neben den infrastrukturellen Abhängigkeiten, die die Abbildung von eGovernment-Prozessen an Hochschulen mittels OSCI erfordert, müssen die entsprechenden Prozesse nach vorausgegangener Analyse zusammen mit den jeweilig anfallenden Daten in einem OSCI-Inhaltsdaten-XML-Schema spezifiziert und modelliert werden. Ausgehend von diesem XML-Schema können dann eGovernment-Anwendungen für diese Prozesse implementiert bzw. ggf. angepasst werden.

#### **4 eGovernment für Hochschulen und Studierende**

Die Umsetzbarkeit von Verwaltungsverfahren mittels eGovernment-Standards wird u.a. an ausgewählten Beispielen im Projekt SeDiGov im Rahmen des vom Kultusministerium geförderten IuK-Kompetenzzentrums an der Hochschule Harz untersucht. Dabei wird entsprechend der Kategorisierung aus Abschnitt 1 vorgegangen.



#### 4.1 Klassifikation von eGovernment-Verfahren

Die Klassifikation von eGovernment an der Hochschule Harz setzt eine Identifikation von möglichen eGovernment-Kandidaten sowie deren Prozessanalyse voraus. Im Rahmen des IuK-Kompetenzzentrums wurden als Vertreter für die oben genannten drei Kategorien dann Anwendungen wie „interne Prüfungsverwaltung“, „Fördermittelmanagement mit externen Stellen“ und „Bologna-Prozess“ als relevante Prozesse ausgewählt, um entsprechende Prozessanalyse und dann Varianten für deren konzeptionelle eGovernment-Umsetzungen zu erproben. Erste Analyse-Ergebnisse wurden mittels UML-Modellierung dargestellt. In einer zweiten Phase werden Konzepte erarbeitet, um ausgewählte Verfahren, insbesondere aus dem Bereich des Bologna-Prozesses mit Umsetzungskonzepten auf Basis von eGovernment-Standards wie OSCI zu versehen.

#### 4.2 eGovernment für den Bologna-Prozess – das XStudy-Framework auf Basis von OSCI

Ein Verwaltungsverfahren, welches sich zur Überführung in eine eGovernment-Anwendung unter Effizienz- und Sicherheitsgesichtspunkten anbietet, stellen die Austausch- und Anrechnungsvorgänge im Rahmen des Bologna-Prozesses für Prüfungsleistungen und Credit-Points dar, die ein Student an einer Gasthochschule erwirbt und dann an der Heimathochschule mittels ein Transcript-of-Records (TOR) zur Anerkennung vorlegt (vgl. Abschnitt 2). Zur Ablösung des papiergebundenen Datentransportes kann unter Effizienz-, Sicherheits- und Standardisierungsgesichtspunkten ein OSCI-basiertes Inhaltsdaten-Format eingesetzt werden.

Ausgehend von entsprechenden UML-Sequenzdiagrammen der Vorgänge sowie den darin abgebildeten Akteuren und benötigten Daten, die diese miteinander austauschen, wurde ein XStudy-Modell-Framework entworfen, s. Abb. 3, welches in Hinblick auf den OSCI-Infrastrukturstandard eine Bündelung von XML-Inhaltsdatenschemata darstellt.

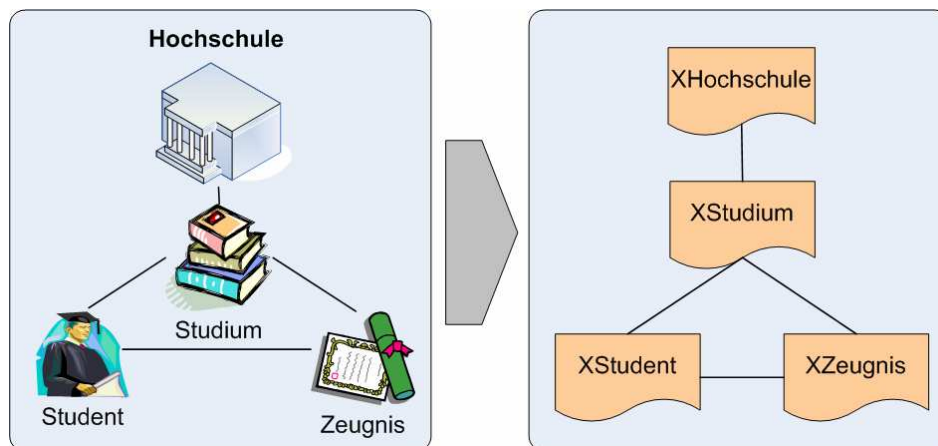
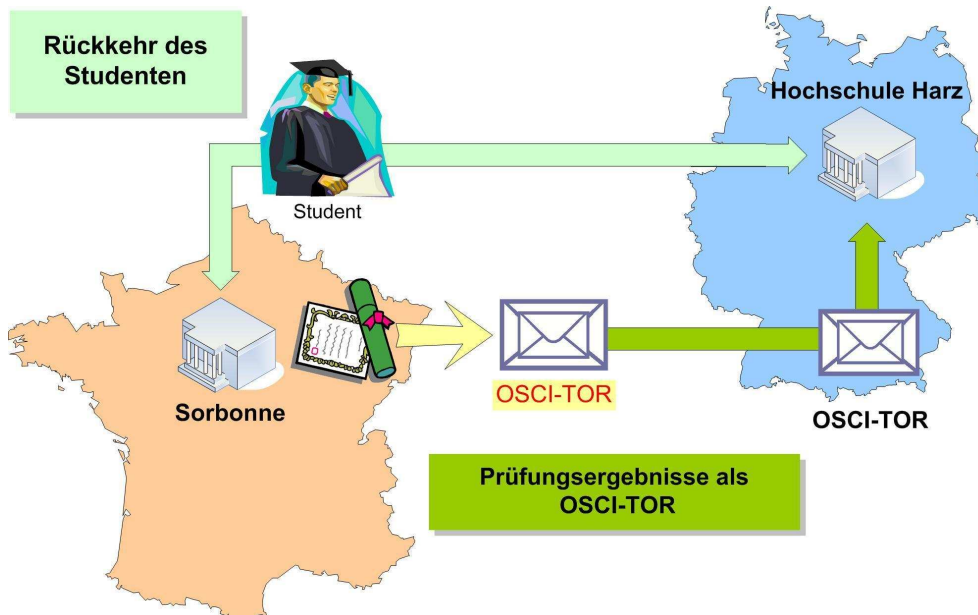


Abb. 3: XStudy-Framework

Grundlage dieses Modells sind die vier XML-Schemata „XHochschule“, „XStudium“, „XStudent“ und „XZeugnis“, welche miteinander in Beziehung stehen. So beschreibt ein XHochschul-Dokument die relevanten übergeordneten Hochschulstrukturen (wie Fachbereiche). Verschiedene Studiengänge mit den jeweiligen obligatorischen und optionalen Anteilen werden als Studien- und Prüfungsordnungen mittels XStudium dargestellt. Der Studienverlauf eines Studierenden dieser Hochschule, der bestimmte Studienleistungen absolviert, wird über ein XStudent-Dokument als „virtuelles Studienbuch“ beschrieben. Der Studierende kann (z.B. bei Auslandsaufenthalten) für einzelne Leistungen dann Credits und



Leistungspunkte des Studienangebots erwerben, die ihm dann in einem Prüfungs-Dokument bestätigt werden (z.B. als TOR).



Um diesen Prozess mit Hilfe der Xstudy-Datenstrukturen via OSCI elektronisch abbilden zu können, müssen TOR-Nachrichten – ebenfalls als XML-Schema – definiert werden, die die angesprochenen Datenstrukturen der Hochschulen miteinander mittels entsprechender prozessorientierter Nachrichten verbinden, so dass damit der TOR-Datenaustausch über eine OSCI-Infrastruktur stattfinden kann. Diese Nachrichten sind über XStudy-Schemata definiert und sind für unterschiedliche Verfahren anpassbar. Für das Beispiel des Bologna-Teilprozesses der Datenabschrift (Übertragung von Credit-Points und Prüfungsleistungen als TOR) wird eine XStudy-Nachricht des Typs TOR/Datenabschrift mit entsprechenden Teilnachrichten festgeschrieben, die als OSCI-TOR-Inhaltsdaten in eine OSCI-Nachricht eingebracht werden können.

<b>Merkmal</b>	<i>XStudy mit OSCI</i>	<i>XStudy ohne OSCI</i>
Integrierte Inhaltsdatenstruktur	+	–
Rechtsverbindliche Zustellung	+	–
Nutzung bestehender Infrastrukturen	+	–

## 5 Resümee und Ausblick

Es wurden Umsetzungsmöglichkeiten für gesichertes und rechtsverbindliches eGovernment an Hochschulen, auch im Bolognaprozess, auf Basis des OSCI-Standards aufgezeigt – mit hohem Effizienzpotential. Weitere entsprechende Untersuchungen anderer Verfahren folgen (z.B. für Notenübermittlungen an Prüfungsämter, Bestellungen oder Deputatsabrechnungen).

## Literatur:

- [0] Landesregierung Sachsen-Anhalt: eGovernment-Aktionsplan für die Landesverwaltung Sachsen-Anhalt, Version 0.1 zur Veröffentlichung, MI LSA, Magdeburg, 17. August 2004
- [1] OSCI-Leitstelle: OSCI-Transportspezifik.1.2, OSCI-Leitstelle, Bremen, Juni 2002
- [2] Strack, Karich: OSCI als Basis für vertrauenswürdigen eGovernment - auch für Hochschulen?, LDVK-Tagung, Hochschule Harz, Wernigerode, 01/2006
- [3] Strack, Karich: BeGovSAH – Begleitforschung zur Umsetzung des eGovernment-Aktionsplans in Sachsen-Anhalt, Tagung „IT-Sicherheit“ der GI, Magdeburg, 02/2006
- [4] Bundesministerium des Innern: Verordnung zur Durchführung von regelmäßigen Datenübermittlungen zwischen Meldebehörden verschiedener Länder (BMeldDÜV1 2005), BGBl I 2005
- [5] MEDIA@Komm, mediakomm.difu.de: Pilotprojekt zu Government in Deutschland, 1999-2003
- [6] Strack, H.: MEDIA@Komm – the Pilot Project for E-Government and Ecommerce in Germany, Tagungsband, ISSE 2001, Information Security Solutions Europe, London, 2001
- [7] Biernath, J.: Ministry of Finance of Saxony-Anhalt; An ‘electronic butler’ for the Structural Funds in Saxony-Anhalt, Zeitschrift "inforegio", Hrsg. Europ. Kommission, Ausgabe Nr. 16, Mai 2005
- [8] Alesi, Bürger, Kehm, Teichler: Stand der Einführung von Bachelor- und Master-Studiengängen in Bologna-Prozess sowie in ausgewählten Ländern Europas im Vergleich zu Deutschland; Wissenschaftliches Zentrum für Berufs- und Hochschulforschung – Universität Kassel im Auftrag des Bundesministeriums für Bildung und Forschung, Endbericht, Februar 2005
- [9] Bundesministerium für Bildung u. Forschung: Der Bologna Prozess, Internetveröffentl., 2006
- [10] Hochschulrektorenkonf.: Internat. Anerkennung, Service-Stelle Bologna, Internetveröffentl., 2006
- [11] Hochschulrektorenkonferenz: Individuelle Studienwege und Qualifikationsprofile, Service-Stelle Bologna, Internetveröffentl., 2006
- [12] Bundesministerium für Bildung und Forschung, Sekretariat der Ständigen Konferenz der Kultusminister der Länder, et. al.: Realisierung der Ziele des Bologna-Prozesses, Nationaler Bericht 2004 für Deutschland unter Mitwirkung von Hochschulrektorenkonferenz et. al., 2004
- [13] Assenmacher, Hoffmann, Kurth, Leupold, Milius, Nühlen, Schöps, Segelken, Stehfest: Handbuch Einführung eines Leistungspunktsystems im Fachhochschulverbund Sachsen-Anhalt, Hrsg. Hochschule Anhalt (FH), mit Unterstützung des Kultusministeriums des Landes Sachsen-Anhalt, in Kooperation mit den Hochschulen des Landes Sachsen-Anhalt, Juni 2002

## Dezentrales Bewertungskonzept für Web Services

Nico Brehm, Phone: ++49.441-798-4474, Fax: ++49.441-798-4472,  
E-Mail: nico.brehm@uni-oldenburg.de, Carl von Ossietzky Universität Oldenburg, Ammerländer Heerstrasse 114-118, D-26129 Oldenburg

**Abstract.** Das Service-orientierte Architekturparadigma zielt darauf ab, die Ressourcen eines Systems in Form von Diensten on-demand bereitzustellen. Moderne Softwaresysteme berücksichtigen dieses Modell bereits und die Vielzahl der Standardisierungsansätze im Bereich der Web Service-Technologie zur Umsetzung von SOA ist zukunftsweisend. Jedoch ist die Vision der automatischen Generierung von Programmen zur Kommunikation mit Web Services und die automatische Suche und Entscheidung über die Nutzung von konkurrierenden Diensten dem Sicherheitsproblem ausgesetzt, dass Dienstnutzern keine Informationen über die Vertrauenswürdigkeit von fremden Dienst Anbietern zur Verfügung stehen. Der vorliegende Ansatz beschreibt ein Sicherheitsprotokoll (DBP1) durch dessen Anwendung Bewertungen über Dienste dezentral von den Anbietern selbst verwaltet werden können. Das Sicherheitsprotokoll zielt auf die Verhinderung von Manipulationen durch die Dienstanbieter ab.

**Keywords:** Trust, PKI, Web Service, SOA, Reputation

### 1 Einleitung

Bezugnehmend auf die Anzahl laufender Aktivitäten zur Standardisierung der semantischen Beschreibung von Web Services oder als Alternative der Einführung intelligenter Suchmechanismen für Web Services ist davon auszugehen, dass die Integration von Softwarekomponenten externer bzw. fremder Anbieter in zukünftigen Softwarearchitekturen eine maßgebende Eigenschaft sein wird. Jedoch ist die Realisierung der daraus resultierenden Vision einer herstellerübergreifenden Kopplung von Softwarekomponenten auf der Basis verteilter Systemarchitekturen einer Vielzahl von Risiken ausgesetzt, deren gemeinsamer Ursprung in der Schwierigkeit zur Herstellung einer nachvollziehbaren Vertrauenswürdigkeit und Verlässlichkeit von Diensten und Anbietern liegt. Zur Beantwortung dieser offenen Fragestellung existieren bereits Modelle und Frameworks zur Aushandlung von Vertrauensbeziehungen in offenen Netzwerkarchitekturen wie Grids und Peer-to-Peer-Netzwerken. Zentralisierte Architekturen zur Sammlung von Bewertungen über Dienste und Dienstanbieter haben dabei aus der Sicht von offenen Netzen den entscheidenden Nachteil, dass eine dauerhafte Abhängigkeit von der Bewertungszentrale besteht. Bei der Einführung und Umsetzung nutzungsbasierter Geschäftsmodelle für Web Services, wobei Konkurrenzbeziehungen zwischen Web Service-Anbietern charakteristisch zur Konstruktion eines Komponentenmarkts sind, senken zentrale Instanzen die Akzeptanz des Gesamtsystems und bilden aus der Sicht der Hersteller von Komponenten eine Hemmschwelle zum Einstieg in einen neuen Absatzmarkt. Im folgenden wird ein dezentrales Konzept zur Bewertung von Web Services beschrieben, das auf einem speziell entwickelten PKI-Protokoll basiert.

### 2 Dienstbewertungsprotokoll (DBP1)

Im Folgenden wird das Dienstbewertungsprotokoll (DBP1) zur Bewertung von Diensten beschrieben. Ziel der Anwendung des Protokolls ist die gesicherte Abgabe von Bewertungen von Dienstnutzern im Bezug auf Dienste, die von Providern angeboten werden. Um auf eine zentrale Stelle zur Speicherung aller Bewertungen aller Dienstanutzer im Netzwerk zu

verzichten, sollen die Bewertungsdaten von den bewerteten Providern selbst verwaltet werden. Durch die Anwendung entsprechender Sicherheitsmechanismen innerhalb des Protokolls sollen Manipulationen der Bewertungen durch die Provider verhindert werden. Bewertungsdaten werden dabei über vertrauenswürdige Vermittlungsinstanzen temporär aufbereitet. Manipulationsmöglichkeiten der Dienstnutzer im Bezug auf die Korrektheit der abgegebenen Bewertungen werden nicht berücksichtigt.

## 2.1 Voraussetzungen und verwendete Sicherheitsalgorithmen

Voraussetzung für den gesicherten Ablauf einer Bewertungsabgabe sowie einer gesicherten Aufbereitung der Bewertungsdaten ist das Vorhandensein eines Public-Key-Schlüsselpaares bei allen Beteiligten. Für den Beweis der Echtheit öffentlicher Schlüssel wird vom Vorhandensein entsprechender Zuordnungen zwischen persönlichen Identifikationsdaten und öffentlichen Schlüsseln in Zertifikaten ausgegangen, die von entsprechend autorisierten Zertifizierungsstellen (CA) ausgestellt wurden. Zum Schutz des privaten Schlüssels wird die Abstimmung der sicherheitsspezifischen Eigenschaften von Datenträgern mit dem Anwendungsfall hingewiesen.

### Sicherheitsalgorithmen:

- Asymmetrischer Kryptoalgorithmus zum Verschlüsseln  $e(PK, M)$  und Entschlüsseln  $d((PK, M), SK) = M$ ; ( $PK(T)$  öffentlicher Schlüssel und  $SK(T)$  privater Schlüssel des Teilnehmers  $T$ , Nachricht  $M$ )
- Digitales Signaturverfahren mit Signatur  $SIG = s(SK(T), M)$  und Verifikation mit dem öffentlichen Schlüssel  $PK(T)$  durch Prüfung von  $M = v(SIG, PK(T))$
- (Pseudo-) Zufallszahlengenerator mit Output  $r_i$ , wobei  $i$  für die erzeugende Instanz der (Pseudo-) Zufallszahl steht
- Message Digest  $h$

## 2.2 Instanzen und Protokollelemente

Folgende Instanzen müssen zur Beschreibung der Funktionsweise des Protokolls berücksichtigt werden.

- Dienstnutzer  $DN$ , der eine Bewertung  $B_D$  abgibt, die sich auf einen Dienst  $D$  bezieht
- Provider  $P$ , der eine Bewertung  $B_D$  über seinen eigenen Dienst  $D$  empfängt
- Bewertungsaufbereitungsstelle  $BAS$ , die eine Zusammenfassung  $Sum(D)$  über alle bisherigen Bewertungen bezüglich eines Dienstes  $D$  erstellt ein Archiv  $ARC(BAS)$  über alle von ihr selbst berücksichtigten Bewertungen  $B_i$  verwaltet.

### Aufbau einer Bewertung:

Eine Bewertung  $B_D$  besteht aus zwei Teilen, einem Identifikationsteil (Header) und einem Bewertungsteil (Body) in dem sich der Bewertungsinhalt  $BI$  in verschlüsselter Form befindet. Das Header-Element einer Bewertung besteht aus folgenden Teilelementen:

- Identifikation der Bewertung als (Pseudo-) Zufallszahl  $r_{DN}$ , die vom Dienstnutzer bei der Erstellung einer Bewertung erzeugt wird.

N. Brehm

- Zertifikat des Providers  $C(P)$ , der den zu bewertenden Dienst anbietet
- Zertifikat des Dienstnutzers  $C(DN)$ , der die Bewertung erstellt hat
- Zertifikat der Bewertungsaufbereitungsstelle  $C(BAS)$ , für die der Bewertungsinhalt im Body-Element der Bewertung verschlüsselt ist.
- Identifikation des Dienstes  $Id(D)$  auf den sich die Bewertung bezieht

Das Bodyelement einer Bewertung beinhaltet den Bewertungsinhalt  $BI$ , der für die BSA verschlüsselt ist  $e(PK(BSA), BI)$ .

### **Aufbau einer Bewertungszusammenfassung:**

Eine Bewertungszusammenfassung besteht aus zwei Hauptelementen, einem Verifikationsteil (Header) und einem Bewertungsinhaltsteil (Body). Das Headerelement setzt sich aus folgenden Teilelementen zusammen:

- Zertifikat der  $BAS$ , von der die Bewertungszusammenfassung erstellt wurde  $C(BAS)$
- Signatur des Bewertungsinhalts, die zur Überprüfung der Echtheit des Bewertungsinhalts dient  $s(SK(BAS), BI)$

Das Bodyelement setzt sich aus folgenden Teilelementen zusammen:

- Identifikation des Dienstes  $Id(D)$ , die zur Herstellung eines Zusammenhangs zwischen den eigentlichen Zusammenfassungsinhaltsdaten und dem bewerteten Dienst verwendet wird
- Identifikation des Providers  $Id(P)$ , die zur Herstellung eines Zusammenhangs zwischen Provider und Dienst notwendig ist.
- Zusammenfassungsinhaltsdaten  $ZID$ , die die aufbereitete Gesamtbewertung aller berücksichtigten Bewertungen von Dienstnutzern im Klartext repräsentiert.

## **2.3 Funktionsweise**

Zur Beschreibung der Funktionsweise wird zwischen der Abgabe einer Bewertung, der Erstellung einer Bewertungszusammenfassung und der Abfrage einer Bewertungszusammenfassung unterschieden.

### Abgabe einer Bewertung:

1. DN an P:  $Init(Id(D))$
2. P an DN:  $C(BAS)$
3. DN an P:  $B_D, s(SK(DN), h(B_D))$

### Erstellung einer Bewertungszusammenfassung:

1. P an BAS:  $B_D, s(SK(DN), h(B_D)), Sum(D)_t$
2. BAS an P:  $Sum(D)_{t+1}$

### Abfrage einer Bewertungszusammenfassung:

1. DN an P:  $Get(Id(D))$
2. P an DN:  $Sum(D)_t$

### **3 Verwandte Arbeiten und Ausblick**

Es existiert eine Vielzahl von reputationsbasierten Vertrauensmodellen, die auf der Abgabe von Bewertungen basieren. Einem Großteil dieser Ansätze liegt eine zentralisierte Strategie zugrunde bzw. es wurde bei diesen Arbeiten keine dezentralisierte Variante beschrieben. Als zentralisierte Strategie wird hier eine Strategie zur Verwaltung von Bewertungen bezeichnet, deren zugrunde liegendes Modell auf einem oder mehreren zentralen Elementen basiert, die permanent Bewertungen anderer Teilnehmer vorhalten. Weiterhin existieren Bewertungsmodelle im Bereich Peer-to-Peer Netze und Agentensysteme, die keine zentrale Instanz beinhalten, sondern davon ausgehen, dass Bewertungen über Peers vor der Dienstnutzung von benachbarten Peers eingeholt werden. Zumeist steigt dabei mit zunehmender Anzahl befragter Nachbarknoten die Genauigkeit der Reputationsmessung. Der Unterschied zum vorgestellten Modell ist, dass dabei einem anfragenden Client die aktuellsten Daten in umfassender Form und ohne die Kontaktaufnahme zu anderen Clientknoten direkt vom betroffenen Provider bereitgestellt werden kann. Die Verwaltungskosten zur Speicherung von Bewertungen können auf die Provider von Web Services übertragen werden. Bewertungsaufbereitungsstellen können in beliebiger Anzahl im Netzwerk existieren und unter Berücksichtigung verschiedenartiger Geschäftsmodelle die Erstellung von Bewertungszusammenfassungen als Dienstleistung anbieten.

# **Informatik / Wissen für die Zukunft**





# Management von Serviceprozessen im HCC Magdeburg auf Basis des ITIL-Frameworks

Prof. Dr. Claus Rautenstrauch, Phone ++49 391 67 1 83 85, Fax ++49 391 67 1 12 16,  
E-mail: rauten@iti.cs.uni-magdeburg.de, Otto von Guericke Universität, Universitätsplatz 2, 39106 Magdeburg

Prof. Dr. Hans-Jürgen Scheruhn, Phone: ++49 3943-659-379, Fax: ++49.3943-659-399,  
E-Mail: hscheruhn@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

Dipl.Wirtsch.-Inf. Stefan Weidner, Phone: ++49 391 67 1 85 27, Fax: ++49 391 67 1 12 16,  
E-Mail: weidner@hcc.uni-magdeburg.de, Otto von Guericke Universität

**Abstract:** ITIL hat bereits in vielen IT-Service-Bereichen Einzug gehalten. In dem folgenden Beitrag zeigen die Autoren zunächst, dass die typischen Abläufe in einem SAP Hochschulkompetenzzentrum (HCC) auf Basis des ITIL-Framework abbildbar sind. Im zweiten Schritt sind diese im Kontext des Geschäftsprozessmanagements umzusetzen. Die Definition der strategischen Ziele erfolgt über eine ITIL-spezifische Balanced Scorecard. Die daraus resultierende modellbasierte Implementierung der Sollprozesse nutzt den SAP Solution Manager bzw. SAP ERP ECC. SAP Business Workflow / WebFlow und geeignete SAP Business Objects sollen sowohl die Prozessführung als auch das Monitoring der ITIL-basierenden Sollprozesse unterstützen.

**Keywords:** Service-Prozesse, SAP Hochschulkompetenzzentrum Magdeburg, Education Service Providing, HCC-ITIL-Referenzmodell, ARIS, Balanced Scorecard, SAP Solution Manager, SAP ERP ECC, Modellbasierte Implementierung, SAP Business Workflow/WebFlow, SAP Business Objects

## 1 ITIL-Service-Prozesse des HCC Magdeburg

### 1.1 Das HCC Magdeburg

Die Otto-von-Guericke-Universität Magdeburg betreibt zusammen mit der SAP AG, Hewlett Packard und T-Systems das weltweit größte SAP Hochschulkompetenzzentrum und bietet Hochschulen und anderen öffentlichen Bildungseinrichtungen Zugang zu SAP-Softwarelösungen. Das HCC dient dabei ausschließlich Lehrzwecken, dem Zweck der internen Aus- und Weiterbildung von Dozenten sowie der Erstellung von wissenschaftlichen Arbeiten (Seminar-, Diplom-, Master- und Doktorarbeiten) im Umfeld von Enterprise Resource Planning (ERP)-Systemen. Das schnelle Wachstum des HCC Magdeburg auf nunmehr 130 angeschlossene Institutionen mit über 1300 Dozenten und etwa 40.000 Studierenden machte die Optimierung der gewachsenen IT-Prozesse notwendig [HCC06].

### 1.2 ITIL im HCC Magdeburg

ITIL ist die Abkürzung für den im Auftrag der britischen Regierung durch die Central Computer and Telecommunications Agency<sup>1</sup> (CCTA) entwickelten Leitfaden "IT Infrastructure Library", der seit 1989 vom Office of Government Commerce<sup>2</sup> (OGC) veröffentlicht wird. ITIL stellt ein Framework aus Best Practices dar, anhand derer IT-

---

<sup>1</sup> <http://www.ccta.gov.uk/>

<sup>2</sup> <http://www.ogc.gov.uk/index.asp?id=2261>

Prozesse definiert, optimiert und gesichert werden können.

Ein wichtiges Merkmal des ITIL-Frameworks ist dessen Prozessorientierung [Rup02]. Fokus der Betrachtung im Kontext der HCC-Geschäftsprozesse bilden in diesem Beitrag ausgewählte HCC-Teilprozesse aus den ITIL-Bereichen „Service Support“ auf operativer Ebene [ITIL00] und „Service Delivery“ auf taktischer Ebene [ITIL01]. Strategisches Hauptziel ist der nachhaltige Ausbau stabiler HCC-Services.

Die Implementierung der ITIL-Serviceprozesse erfolgt im Rahmen eines Prozessmanagementprojekts [OPM06]. Der von ITIL-Nutzern oft als zu niedrig eingeschätzte Detaillierungsgrad des ITIL-Regelwerkes [Myr04] macht es erforderlich, dass auch die Autoren dieses Beitrags eine konkrete Umsetzung und Ausgestaltung vornehmen und somit ein HCC-ITIL-Referenzmodell (Englisch / Deutsch) für Education Service Providing (ESP) schaffen. ESP bedeutet Application Service Providing (ASP) für SAP, erweitert um die speziellen Anforderungen im Ausbildungsbereich [HCC06].

Die Akteure sind gemäß ITIL neben den Leistungserbringern der IT-Services die Kunden eines HCC (alle angeschlossenen Institutionen wie z.B. Hochschulen) bzw. die Anwender (z.B. Dozenten) als Leistungsabnehmer. Als Lieferanten sollen die Ressourcen des HCC, z.B. SAP (Lizenzen), HP (Hardware), Otto-von-Guericke-Universität Magdeburg (Personal, Infrastruktur) und T-Systems (Know-how) betrachtet werden (vgl. Abb.1 und Abb.2).

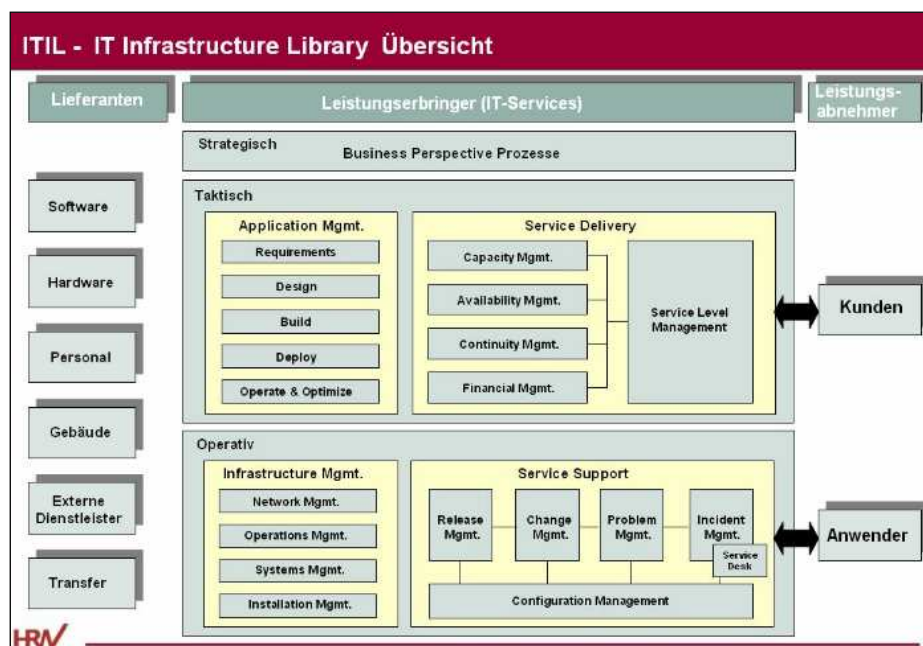


Abb.1: ITIL-Framework [ vgl. Hoch03]

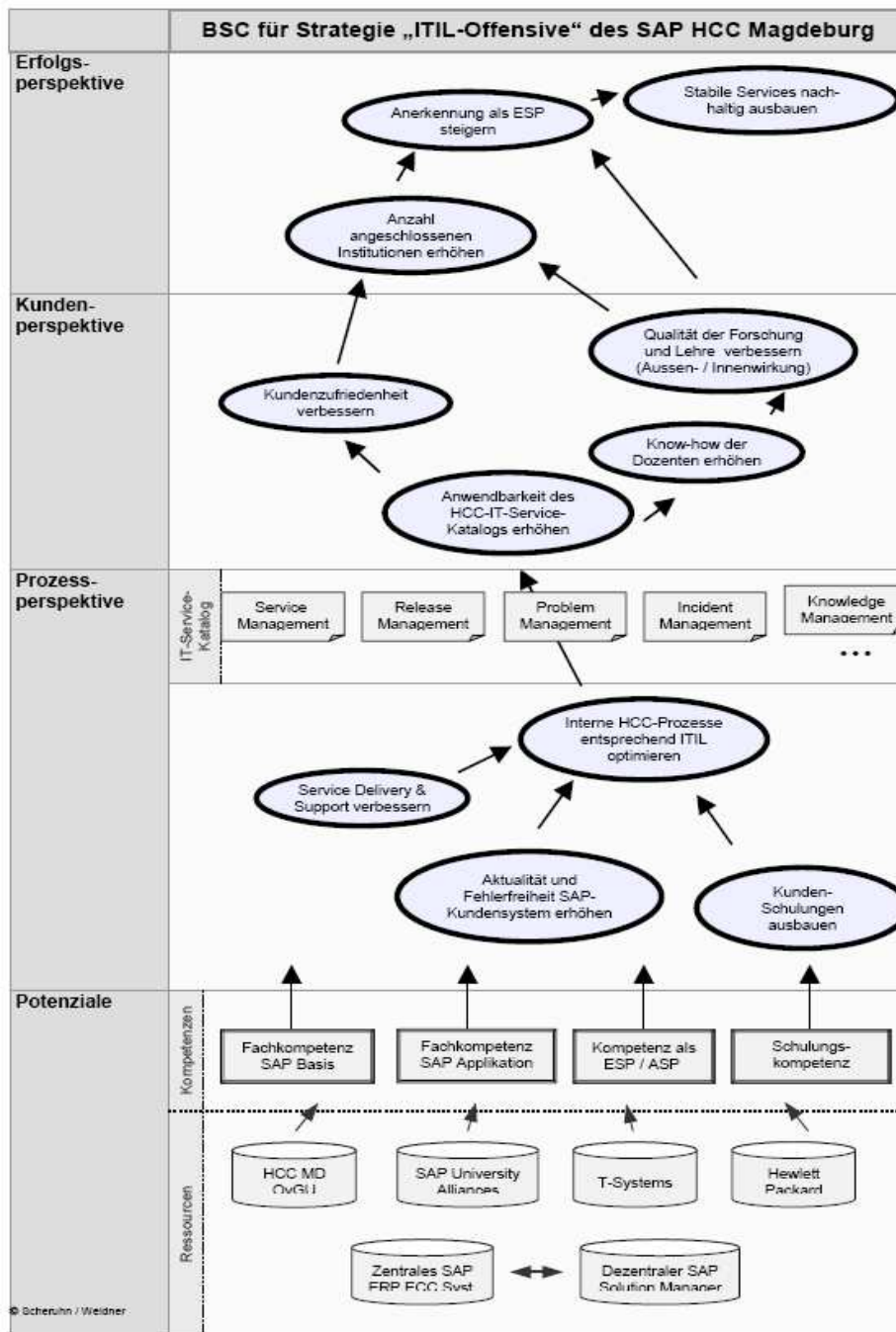


Abb.2: Balanced Scorecard für die Strategie „ITIL-Offensive des HCC Magdeburg“

Die oben erwähnten strategischen Ziele [ITIL04] der HCC lassen sich auf Basis der Balanced Scorecard-Methode als Ursache-Wirkungskette darstellen (vgl. Abb.2.). Bei der Balanced Scorecard (BSC) handelt es sich um eine in den 90er Jahren von Robert Kaplan und David Norton entwickelte, ganzheitliche Managementmethode. Dabei werden aus einer langfristigen Vision und den daraus abgeleiteten Strategien – hier einer ITIL-Offensive - heraus Zielvorgaben vier verschiedener Betrachtungsperspektiven (Erfolg, Kunden, interne

Prozesse und Potenzial) kommunizierbar, messbar und letztlich auch umsetzbar gemacht. Als Führungsinstrument eingesetzt, verhindert die BSC die zu einseitige interne Ausrichtung auf monetäre Ziele und spannt einen Bezugsrahmen für monetäre und nicht-monetäre Ziele gleichermaßen. Für den Grad der Zielerreichung sieht die BSC-Methode so genannte Key Performance Indicators (KPI) als Kennzahlen vor, die je eine Ausprägung für den Soll-Zustand bzw. Ist-Zustand besitzen [Kap97].

Die an die HCC „outgesourcten“ Serviceprozesse sollen die Kernprozesse – d.h. Forschung und Lehre - der angeschlossenen Kunden – d.h. der Hochschulen - unterstützen [Rau06]. Die strategischen Ziele der Kundenperspektive der HCC (vgl. Abb.2) korrespondieren dabei mit den dort dargestellten strategischen Zielen der Prozessperspektive des Kunden.

Neben einer Erfolgs- bzw. Kundenorientierung verdeutlicht die hier exemplarisch betrachtete BSC für das HCC Magdeburg den Ursache-Wirkungs-Zusammenhang mit den Serviceprozessen und den dazu benötigten Potentialen wie Kompetenzen und Ressourcen des HCC Magdeburg (vgl. Abb. 2). Bei den Serviceprozessen wird eine Auswirkung auf die Anwendbarkeit durch die Hochschulen – direkte Kunden im Kontext der BSC – betrachtet. Bei den Kompetenzen stehen z.B. Fachkompetenzen (SAP Basis/SAP Applikationen oder Schulungskompetenz) im Mittelpunkt.

Die Ressourcen umfassen die Mitarbeiter des HCC Magdeburg, deren Hardware-, Software- und Infrastrukturlieferanten und ein zentrales SAP ERP ECC-System bzw. dezentrale SAP Solution Manager-Systeme als Implementierungsplattform.

## 2 ITIL-Prozessmanagement mit SAP-Systemen

In der Literatur findet sich keine eindeutige Definition für den Begriff „Prozessmanagement“. Dieser kann als Regelkreis [Str05] oder als kontinuierlich zu durchlaufender Prozess-Lebenszyklus betrachtet werden, der aus den Phasen

- Zieldefinition (ITIL: Requirements)
- Prozessdefinition mit Prozessführung / Workflow (ITIL: Design, Build und Deploy)
- Prozessausführung (ITIL: Operate)
- Prozess-Monitoring / Controlling (ITIL: Optimize)

besteht [Sche04]. Den einzelnen Phasen lassen sich auch die 6 ITIL-Teilprozesse des „Application Management“ (s.o. und vgl. Abb. 1) zuordnen.

Während die Umsetzung der letzten drei Phasen z.B. über die SAP ECC -Module PM (Plant Maintenance), CS (Customer Service) bzw. PS (Projektsystem) erfolgen kann [SAP05], ist es das Ziel dieses Beitrages, insbesondere die möglichen Vorteile des SAP Solution Managers gegenüber einem oder im Zusammenspiel mit einem SAP ERP ECC-System zu untersuchen. Die Integration der o.g. drei Phasen soll in jedem Fall mittels des SAP Business Workflow/WebFlow erfolgen.

Abb. 1. definiert eine mögliche ITIL-Geschäftsprozessstruktur und gibt bereits die grobe Richtung des Dienstleistungs-, Finanz- bzw. Informationsflusses vor.

Eine horizontale Segmentierung (z.B. in einen Incident-Management, Problem-Management bzw. Installation-Management-Teilprozess gemäß. Abb. 1) und eine vertikale Segmentierung

(Detaillierung z.B. des Teilprozesses „Incident-Management“ in die „Incident-Entdeckung“ und „Klassifizierung“ bis „Incident-Abschluss“) gemäß Abb. 3 ist Voraussetzung für eine strukturierte Erfassung des Ist-Zustands und eine zielführende Vorgehensweise innerhalb der nachfolgenden Phasen [Sche04].

Unerlässliche Hilfsmittel zur Definition der horizontalen aber auch vertikalen Segmentierung sind existierende ITIL-Referenzmodelle, z.B [Pul05] oder [IDS06].

Im Folgenden betrachten die Autoren im Kontext der vertikalen Segmentierung unterhalb der ITIL-Teilprozesse wie z.B. Incident-Management nur drei weitere Ebenen (vgl. Abb. 3) Die Ebene der Basis-Aktivitäten (z.B. „Incident-Entdeckung“), die aggregierte Übersichtsebene (z.B. „Benachrichtigen des Support-Teams“) und die darunterliegende Detailebene (z. B. „Incident aufzeichnen“) welche bereits die operativen Tätigkeiten auf der Ebene von Arbeitsanweisungen beschreibt, die im ITIL-Framework nicht weiter spezifiziert sind. Die hierfür benötigten Informationen mußten durch Interviews mit den Mitarbeitern des HCC gewonnen werden.

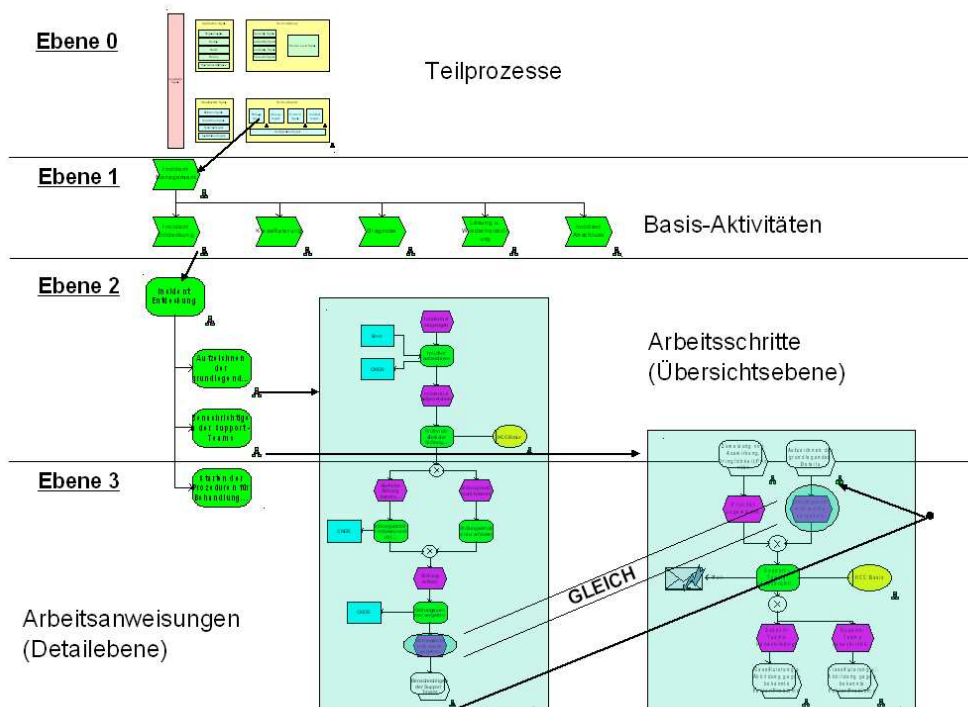


Abb.3: Horizontale und vertikale Segmentierung im Vier-Ebenen-Modell

Wichtigster struktureller Bestandteil von Geschäftsprozessen sind Aufgaben bzw. Funktionen. Zur vereinfachten Darstellung der horizontalen und vertikalen Segmentierung von Geschäftsprozessen eignet sich daher insbesondere Wertschöpfungsketten bzw. hierarchische Funktionsbäume. Auf unterster Ebene finden erweiterte ereignisgesteuerte Prozessketten (eEPKs) Verwendung, welche auch die Arbeitsabläufe (Workflows) in den SAP-Systemen spezifizieren. Bei diesen Prozessketten handelt es sich um eine grafische

Darstellungsform für Geschäftsprozesse, die ein wesentlicher Bestandteil des ARIS-Konzepts<sup>3</sup> der IDS Scheer AG<sup>4</sup> ist.

Auf dieser Ebenen kann auch ein Sprung in einen anderen Arbeitsschritt aus einer anderen Basisaktivität erfolgen.

### 3 Umsetzung Incident Management

Da das HCC Magdeburg in den untersuchten Service-Support-Bereichen (Incident-, Problem- und Release-Management) alle wesentlichen Teilprozesse, Basisaktivitäten und Arbeitsschritte gemäß ITIL abdeckt, kann ein mögliches Verbesserungspotenzial im Ist-Zustand des HCC Magdeburg erst auf der Ebene der Arbeitsanweisungen (vgl. Abb.3) identifiziert werden. Verbesserungsfähig (qualitativ und zeitlich) ist hier insbesondere die vom Anwender (z.B. Dozent) per email-Kontakt ausgelöste Störmeldung (Incident-Management), welche im HCC einen typischen Medienbruch erzeugt. Auch die weitere Abwicklung der Service-Support-Prozesse erfolgt im HCC nur teilweise integriert auf Basis der vorhandenen Ressourcen wie den SAP-Systemen des HCC oder den Kompetenzen des HCC (vgl. Potenzialperspektive in Abb.2).

Für die integrierte Umsetzung der letzten 3 Phasen des o.g. Prozessmanagements im Soll-Zustand mittels SAP Business Workflow/WebFlow müssen geeignete SAP Business Objekte im mySAP ERP ECC bzw. mySAP Solution Manager identifiziert werden. Dazu wurden vier verschiedene Alternativen näher untersucht:

- Servicemeldung / BUS\_2080 (mySAP ERP ECC)
- Allgemeine Meldung / BUS\_7051 (mySAP ERP ECC)
- Instandhaltungsmeldung / BUS\_2038 (mySAP ERP ECC)
- Servicevorgang / BUS2000116 (my SAP SolutionManager)

Eine nach verschiedenen Kriterien gewichtete Analyse [OPM06] liefert letztlich aufgrund der leichter zu automatisierenden Übertragung von Systemdaten einen Vorteil für den nur im Solution Manager verfügbaren „Servicevorgang“ BUS\_2000116, dicht gefolgt von der nur im ERP ECC verfügbaren „Servicemeldung“ BUS\_2038. Die beiden anderen untersuchten Business Objekte verfügen über keine ITIL-konforme Lösungsdatenbank.

Die für den automatisierten Start einer Business Workflow/WebFlow Instanz benötigten Ereignisse (z.B. Störmeldung erzeugt) im HCC können vom Anwender sowohl über das SAP Employee Self Service (ESS) als auch über eine nach den Vorgaben des HCC entwickelte SAP Business Server Page erzeugt werden. Diese sind ebenso wie die als Business Application Programming Interface [Bapi06] genannten Methoden (Arbeitsanweisungen im Kontext von Abb.3) und Attribute (z.B. Störmeldungs-ID) genau einem der untersuchten Business Objekte zugeordnet.

---

<sup>3</sup> Ziel des ARIS-Konzepts (ARIS = Architektur integrierter Informationssysteme) von Prof. A. Scheer ist es, betriebliche Informationssysteme optimal an ihre Anforderungen anzupassen [Scheer92].

<sup>4</sup> <http://www.ids-scheer.de>

## 4 Umsetzung Service Level Management

Der Teilprozess Service Level Management (SLM) bildet die Schnittstelle zum Kunden (z.B. Hochschule). Er gehört im Unterschied zum im vorherigen Abschnitt betrachteten Teilprozess Incident-Management nicht zum Service Support sondern zum Service Delivery (vgl. Abb. 1). Auch hier ergibt sich eine vertikale und horizontale Prozessstruktur gemäß Abb. 3 [ITIL01].

Der zunächst betrachtete Prozess „Kundenerfassung“ lokalisiert sich in dem HCC-ITIL-Referenzmodell als Arbeitsschritt (vgl. Abb.3) über den Pfad Service Level Management / Implementing the SLM Process / Seek agreement, der darauf folgende Prozess „Dozentennachmeldung“ über den Pfad Service Level Management / Ongoing SLM Process / Maintenance of SLAs<sup>5</sup> contracts and OLAs<sup>6</sup>.

Auch hier lässt sich ein mögliches Verbesserungspotenzial erst auf der Ebene der Arbeitsanweisungen im Ist-Zustand identifizieren. Dies sind insbesondere der im HCC als Fax eingehende Datenerfassungsbogen (DEB) sowie der Dozentennachmeldungsbogen (DNB) und deren weitere nicht vollständig integrierte Verarbeitung auf den HCC-eigenen SAP-Systemen.

Eine Umsetzung des Soll-Zustands geschieht auf Basis des führenden dezentralen SAP Solution Managers bzw. zentralen ERP ECC Systems. Außerdem muß ein bidirektionaler Abgleich ausgewählter Berechtigungsobjekte mit dem dezentralen LDAP-Server möglich sein.

Die Eingabe der Kunden- bzw. Dozentendaten soll zunächst über SAP Business Server Pages auf Portal-/Präsentationsebene erfolgen, die Applikationsebene sorgt über das ausgewählte Business -Objekt „Business Partner / BUS\_1006 für einen konsistenten Zugriff auf die eigentliche Datenebene in Form von geeigneten BAPIs. Die anschließende Auswertung und Aggregation der Daten lässt sich über Reports bzw. SAP Business Intelligence Systeme wie mySAP BW oder mySAP SEM realisieren.

Die Implementierung ausgewählter Arbeitsschritte soll mit dem SAP Business Workflow/WebFlow erfolgen

## 5 Zusammenfassung und Ausblick

Das beschriebene Projekt schaffte die Voraussetzungen für eine umfassende Optimierung der Service-Prozesse im SAP Hochschulkompetenzzentrum Magdeburg. Die Formulierung der strategischen Ziele des HCC und die stringente Ausrichtung der Prozessmodelle am ITIL-Defacto-Standard (auf allen Ebenen) offenbarte Potentiale, die nun nach und nach durch dieses und folgende Projekte genutzt werden sollen. Bereits erkennbare Zeiteinsparungen bestätigen die Erwartungen in dieses Projekt und ermutigen das Management zur umfassenden Anpassung aller Serviceprozessgeflechte im HCC.

In einem weiterführenden Projekt soll u. a. das Schulungsmanagement näher betrachtet werden. Dieses gehört zum Service Support und lokalisiert sich im Ist-Zustand des HCC-ITIL-Referenzmodells über Service Desk / Service Request Handling (Schulungsmanagement) / Schulungsplanung bzw. Durchführung.

---

<sup>5</sup> SLA = Service Level Agreement

<sup>6</sup> OLA = Operating Level Agreement

## Quellenangaben

- [Bapi06] BAPI-Benutzerhandbuch CFA-BA, SAP Knowledge Warehouse  
[http://help.sap.com/saphelp\\_46c/helpdata/de/7e/5e141a4a1611d1894c0000e829fbbd/framset.htm](http://help.sap.com/saphelp_46c/helpdata/de/7e/5e141a4a1611d1894c0000e829fbbd/framset.htm) , vom 27.05.2006
- [HCC06] Das HCC-Konzept des SAP-Hochschulkompetenzzentrums der Otto-von-Guerike-Universität Magdeburg, Darstellung vom 10.02.2006  
<http://www.hcc.uni-magdeburg.de/ssi/public/common/hcckonzept.shtml>
- [Hoch03] Hochstein, A.; Hunziker, A.: Serviceorientierte Referenzmodelle des IT-Managements, HMD Heft 232, August 2003
- [IDS06] ARIS-ITIL-Referenzmodell, Webseite der IDS Scheer AG  
[http://www.ids-scheer.de/international/english/81097/faq\\_support\\_de](http://www.ids-scheer.de/international/english/81097/faq_support_de), vom 27.05.2006.
- [ITIL00] Office of Government Commerce: IT Infrastructure Library Series – Service Support, 2000, Version 1.2.
- [ITIL01] Office of Government Commerce : IT Infrastructure Library Series – Service Delivery, 2001, Version 1.1.
- [ITIL04] Office of Government Commerce : Business Perspective: The IS View on Delivering Services to the Business, 2004
- [Kap97] Kaplan, R. & Norton, D.: Balanced Scorecard. 1. Auflage, Schäffer-Poeschel, Stuttgart, 1997.
- [Myr04] Meier, A.; Myrach. T. : IT – Servicemanagement, HMD Praxis der Wirtschaftsinformatik, 237, 2004.
- [OPM06] Projektbericht Online-Prozessmanagement mit mySAP. Hochschule Harz, 2006
- [Rau06] Rautenstrauch, C., Pegnetter, Richard, Scheruhn, H.J.; Weidener, S.: Strategische Ausrichtung eines internationalen Masterprogramms mit dem Schwerpunkt Integrationskompetenz am Beispiel mySAP, Die neue Hochschule DNH Band 47 Heft1/06
- [Rup02] Rupchock, K. : Planning to Implement Service Management. The Stationery Office Books, 2002.
- [SAP05] SAP AG: ITIL Best practise mit SAP, SAP Info, 125, 2005.
- [Sche04] Scheruhn, H.-J.; Reiter, C. & Gomez, J.: Online Process Management with mySAP.  
Conference Transcript CICE 2004, Santa Clara/Cuba
- [Str05] Strahinger, S.: Business Engineering. HMD Praxis der Wirtschaftsinformatik, 241, 2005.



## Zufallszahlen in der Lehre

Prof. Dr. Thomas Morgenstern, Phone: ++49.3943-659-337, Fax: ++49.3943-659-399, E-Mail: tmorgenstern@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

Prof. Dr Ingo Schütt , Phone: ++49.3943-659- 311, Fax: ++49.3943-659-399, E-Mail: ischuett@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

### Abstract

Die Theorie der Zufallszahlen und ihre Erzeugung berühren viele Teilgebiete der Mathematik, der Informatik, des Operations Research und selbst der Philosophie. Das Thema Zufallszahlenerzeugung eignet sich gleichzeitig sehr gut zur Verknüpfung und Verdeutlichung von zentralen Konzepten aus diesen unterschiedlichen Fachdisziplinen, der Wahrscheinlichkeitstheorie und Statistik. Didaktische und methodische Wege zu finden, dieses umfassende und komplizierte Gebiet lehr- und lernbar darzustellen ist daher herausfordernd.

Die praktische Bedeutung der Erzeugung von Zufallszahlen ist groß, wie Anwendungen in der stochastischen Simulation, im wissenschaftlichen Rechnen und in der Kryptologie bestätigen. In der Lehre können theoretische Überlegungen mit experimentellen Studien kombiniert werden und die Simulation als Laborexperiment auf dem Computer methodisch eingesetzt werden.

**Keywords:** uniform random number generation, statistical tests, stochastic simulation, Mathematics education. 2000 Mathematics Subject Classification: 65C10, 62-XX, 97Dxx.

## 1 Einführung

So einfach teilweise die Erzeugung von Pseudozufallszahlen ist, so schwierig ist es, ein geeignetes Maß für die „Zufälligkeit“ der Zahlen zu finden. Die dafür notwendige Mathematik ist für die Lehre an einer Fachhochschule aufgrund des Stoffumfanges, der Tiefe und der Komplexität nur in groben Zügen darstellbar. Hier bietet es sich an, durch mathematische Experimente die Ergebnisse und Erkenntnisse verständlich und begreifbar zu machen.

### 1.1 Bedeutung von Zufallszahlen

Bei der Frage, ob es „wirklichen“ Zufall gibt bzw. ob man ihn gar erzeugen kann, gerät man schnell in das Grenzgebiet von Mathematik und Philosophie. Allerdings sind Methoden, die zufällig erzeugte Zahlen mit bestimmten Eigenschaften verwenden von sehr großer Bedeutung.

Dabei werden sehr viele oder sehr große Zahlen benötigt, die in ausreichender Zahl und Geschwindigkeit nur durch Computer erzeugt werden können. Durch Programme erzeugte Zahlen sind allerdings deterministisch bestimmt. Solche so genannten Pseudozufallszahlen sind Zahlenfolgen, die für bestimmte Anwendungen „zufällig genug“ sind. Was als „zufällig genug“ gilt, wird anhand bestimmter Maße festgelegt, siehe [2], [3], [4], [7], [8], [10].

## 1.2 Strukturierung

Ein solch umfassendes Gebiet zu lehren, erfordert eine besondere Sorgfalt in der Auswahl der Lehrinhalte. Im Folgenden stellen wir exemplarisch ausgewählte Aspekte der Erzeugung von gleichverteilten Zufallszahlen dar, die in verschiedenen Wahl- und Pflichtveranstaltungen im Grund- und Hauptstudium an der Hochschule Harz gelehrt werden. Zielgruppe sind vor allem Studierende der Informatik, der Automatisierungstechnik und des Wirtschaftsingenieurwesens.

## 2 Motivation

Dass die Erzeugung von „guten“ gleichverteilten Pseudozufallszahlen und ihre Beurteilung nicht einfach ist, wird an einem Beispiel demonstriert.

### 2.1 Der Fibonacci-Generator

Der Fibonacci - Generator wird durch den Algorithmus und Startwerte  $z_1, z_2 \in [0,1)$

$$z_{i+1} = z_i + z_{i-1} \bmod 1$$

bestimmt. Wir wählen als Startwerte  $z_1=0.3$  und  $z_2=0.2$  und erhalten für 500 Werte das Histogramm in Abb. 1. Insbesondere beim Punktdiagramm in Abb. 1 fällt auf, dass das entstandene Gitter nicht gleichmäßig bzw. nicht regelmäßig verteilt ist.

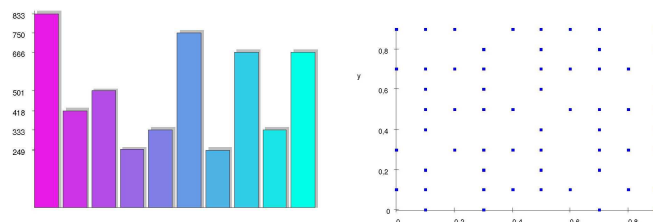


Abb.1: Fibonacci-Generator,  $z_1 = 0.3$   $z_2 = 0.2$

Mit den Startwerten  $z_1=0.32312$  und  $z_2=0.226879$  erhalten wir das Histogramm und das Punktdiagramm in Abb. 2. Das Ergebnis erscheint zumindest deutlich verbessert.

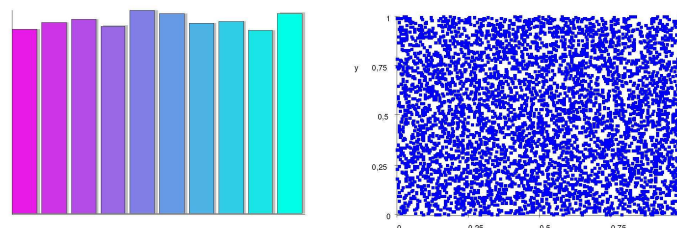


Abb. 2: Fibonacci - Generator,  $z_1 = 0.32312$   $z_2 = 0.226879$

Das dies täuscht und damit auch verdeutlicht, dass bessere Maße für die Güte der Pseudozufallszahlen benötigt werden, zeigen Überlegungen aus [1], denn die beiden Fälle  $z_{i-1} < z_{i+1} < z_i$  und  $z_i < z_{i+1} < z_{i-1}$  können nicht auftreten. Diese Überlegungen legen nun den Verdacht nahe, dass im dreidimensionalen Würfel  $[0,1)^3$  das Punktdiagramm regelmäßige Strukturen aufweist bzw. dass es zu keiner Gleichverteilung der Punkte kommt. Das Abb. 3 zeigt dies in fast unerwarteter Deutlichkeit.

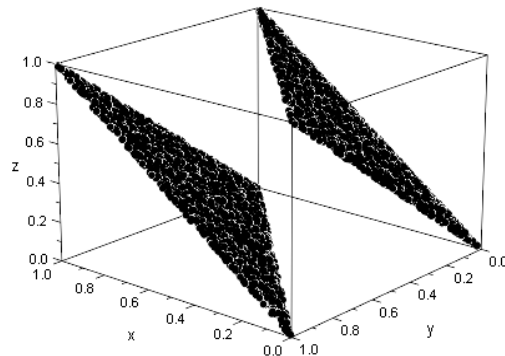


Abb. 3: Fibonacci - Generator, Punktdiagramm

### 3 Grundlagen

Zur Darstellung der mathematischen Grundlagen werden zwei unterschiedliche Gebiete der Mathematik benötigt. Aus der theoretischen Mathematik die Zahlentheorie und aus der angewandten Mathematik die Wahrscheinlichkeitsrechnung und Statistik.

#### 3.1 Zahlentheorie

Die grundlegenden Ideen für lineare Kongruenzgeneratoren zur Erzeugung von Pseudozufallszahlen stammen von D. H. Lehmer [6], Thomson und A. Rotenberg [9].

**Definition 1.** *Linearer Kongruenzgenerator* Sei  $M \subset \mathbb{Z}$ , dann heißt die Abbildung  $t: M \rightarrow M$  ein linearer Kongruenzgenerator, wenn

$$t(x) = (ax + c) \bmod m$$

mit  $a, c, m \in \mathbb{Z}$  gilt. Weiterhin muss gelten:  $m > 0$ ,  $M = \{0, 1, 2, \dots, m-1\}$ ,  $0 \leq a < m$  und  $0 \leq c < m$ .  $m$  heißt Modul,  $a$  heißt Faktor und  $r$  heißt Inkrement.

Ausgehend von einem Startwert (*seed*)  $z_0 \in \mathbb{Z}$  wird die Folge der Pseudozufallszahlen iterativ durch  $z_{n+1} := t(z_n) = az_n + c \bmod m$  erzeugt. Nicht nur in diesem Zusammenhang, sondern z.B. für die Anwendungen in der Kryptologie, werden eine Reihe von zahlentheoretischen Ergebnissen wie die Identität von Bezout, die Kürzungseigenschaft und der Euklidische Algorithmus benötigt und behandelt.

#### 3.2 Wahrscheinlichkeitsrechnung und Statistik

Im Zusammenhang mit Pseudozufallszahlen lassen sich viele Verteilungen einführen: die Multinomialverteilung, die Poissonverteilung, die  $n$ -dimensionale Normalverteilung und Verteilungen im Zusammenhang mit dem Kolmogoroff - Smirnow - Test. In Zusammenhang mit statistischen Tests lassen sich insbes. der Zentrale Grenzwertsatz, die Gammaverteilung und speziell die  $\chi^2$ -Verteilung diskutieren.

Wegen ihrer großen Bedeutung als Kriterium für die Güte von Zufallszahlengeneratoren müssen *Hypothesentests* sehr ausführlich behandelt werden. Da alle Entscheidungen aufgrund von Tests nur mit einer gewissen Wahrscheinlichkeit richtig sein werden, ist die Interpretation der Testergebnisse entscheidend! Die Bedeutung von bestandenen, oder gescheiterten Tests, aber auch die Bedeutung von zu gut bestandenen Tests kann anhand konkreter Beispiele diskutiert werden.

## 4 Methodik

Die theoretischen Aspekte werden in ausführlichen Vorlesungsunterlagen dargestellt, die den Studierenden über das Internet zur Verfügung gestellt werden.

Die Autoren stehen jedoch auf dem Standpunkt, dass die theoretischen Überlegungen an einer Fachhochschule durch experimentelle und praktische Aspekte ergänzt werden müssen. Dazu wurden in dem Computeralgebra-System MuPAD Pro 3.1 alle empirischen Tests aus [5] implementiert. In den Vorlesungsunterlagen werden diese Tests besprochen, die zugrunde liegenden Sätze hergeleitet und die Studierenden zu eigenen Experimenten angeregt.

Als Leistungsnachweis wird im Wahlpflichtbereich die Durchführung und Auswertung einer Simulationsstudie verlangt. Neben vielen anderen praktischen Problemen stößt die Erzeugung und Verarbeitung der Daten dabei regelmäßig an zeitliche- oder kapazitive Grenzen des Computers, wodurch komplexitätstheoretische Aspekte der Berechnungen real erfahrbar werden.

## Literaturverzeichnis

- [1] L. Afflerbach and J. Lehn. *Zufallszahlen und Simulationen*. Teubner, Stuttgart, 1986.
- [2] M. Blum and S. Micali. How to generate cryptographically strong sequences of pseudo-random bits. *SIAM J. Comput.*, 13, 1984.
- [3] O. Goldenreich. Pseudorandomness. *Notices of the AMS*, 46, 1999.
- [4] S. Goldwasser and S. Micali. Probabilistic encryption. *J. Comput. System Sci.*, 28, 1984.
- [5] D. E. Knuth. *The Art of Computer Programming Vol. 2 - Seminumerical Algorithms*. Addison-Wesley, Reading, Massachusetts, Menlo Park, California, New York, 1980.
- [6] D. H. Lehmer. Mathematical methods in large-scale computing units. *Proc. 2nd Sympos. Large-Scale Digital Calculating Machinery*, 1949.
- [7] L. A. Levin. Randomness conservation inequalities: Information and independence in mathematical theories. *Inform. and Control*, 61, 1984.
- [8] M. Li and P. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer - Verlag, New York, 1993.
- [9] A. Rotenberg. A new pseudorandom number generator. *J. Assoc. Comput. Mach.*, 7:75–77, 1960.
- [10] A. C. Yao. Theory and application of trapdoor functions. *rd IEEE Symposium on Foundations of Computer Science*, 1982.

T. Morgenstern

# Uniform Random Binary Floating Point Number Generation

Prof. Dr. Thomas Morgenstern<sup>1</sup>, Phone: ++49.3943-659-337, Fax: ++49.3943-659-399,  
tmorgenstern@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

E-Mail:

## Abstract

Uniform pseudorandom number generators approximate sequences of independent on  $(0,1) \subseteq \mathbf{R}$  uniformly distributed numbers. The output usually lies within the discrete, finite subset of floating point numbers, and therefore can not be random real numbers. Common random number generators produce only a small part of these numbers. Therefore the output can even not be random floating point numbers.

The usual empirical tests are not sensitive enough to detect this. We develop new tests adapted to floating point numbers and successfully construct a new uniform pseudo random binary floating point number generator, by generating the mantissa and exponent independently.

**Keywords:** uniform random number generation, binary floating point numbers, statistical tests, stochastic simulation. 2000 Mathematics Subject Classification: 65C10

## 1 Introduction

Randomised algorithms applied in physics or engineering like Monte Carlo simulation require long sequences of random numbers  $(x_i)_{i \in \mathbf{N}}$ . These sequences have to be samples of independent identically distributed (i.i.d.) random variables  $(X_i)_{i \in \mathbf{N}}$ . Of special interest are on the open interval  $(0,1) \subseteq \mathbf{R}$  uniformly distributed random variables  $X_i \sim U(0,1)$ .

Physical random number generators are often too slow for practical purposes and have to be replaced by software generators. Random numbers produced by algorithmic random number generators are never random but should appear to be random to the uninitiated. We call these generators pseudorandom number generators [1]. They should pass statistical tests of the hypothesis [2]:

$H_0$ : the sequence  $(x_i)_{i \in \mathbf{N}}$  is a sample of i.i.d. random variables.

In fact, no pseudo random number generator can pass all statistical tests. So we may say that bad random number generators are those that fail simple tests, whereas good random number generators fail only complicated tests that are very hard to find and to run [3, 5].

---

<sup>1</sup> This work was established during a research visit at the Institute of Control Systems of the Technical University of Kaiserslautern.

## 1.1 Definitions

To formalize pseudorandom number generators, we follow the definitions in [2-4]:

**Definition 1.** A (pseudo-) random number generator (RNG) is a structure  $(S, \mu, t, O, o)$  where  $S$  is a finite set of states (the state space),  $\mu$  is a probability distribution on  $S$  used to select the initial state (or seed)  $s_0$ ,  $t: S \rightarrow S$  is the transition function,  $O$  is the output space, and  $o: S \rightarrow O$  is the output function.

The state of the RNG evolves according to the recurrence  $s_i = t(s_{i-1})$  for  $i \geq 1$ , and the output at step  $i$  is  $u_i = o(s_i) \in O$ . The output values  $u_0, u_1, u_2, \dots$  are called the random numbers produced by the RNG.

The most widely used RNGs are based on linear recurrence and modular calculations.

**Definition 2.** A multiple recursive generator (MRG) uses:

$$z_i := (a_1 z_{i-1} + \dots + a_k z_{i-k}) \bmod m,$$

where  $m$  and  $k$  are positive integers called the modulus and the order, and the coefficients  $a_1, \dots, a_k$  are in  $\mathbf{Z}_m = [0, \dots, m-1]$ .

The state space is  $S = \mathbf{Z}_m^k$  and the state at step  $i$  is  $s_i = (z_{i-k+1}, \dots, z_i)$ . For  $k=1$  we get the classical (multiplicative) linear congruential generators (LCG)  $z_i := a \cdot z_{i-1} \bmod m$ .

**Example 3.** A widely used good LCG is LCG16807. It was introduced in [7] (and is called RNUN in the IMSL library, LLRANDOM in [8] or URN01 in [9]). It has coefficients  $m = 2^{31} - 1 = 2147483647$  and  $a = 16807$ :

$$z_{n+1} := 16807 \cdot z_n \bmod 2147483647. \quad (1)$$

For values in the interval  $O = (0,1)$  one usually uses the output function  $o$ :

$$u_i = o(z_i) := z_i/m.$$

More complex output functions are of the form:

$$u_i := \sum_{j=1}^l z_{l+i+j-1} \cdot m^{-j}$$

with  $l \leq k$ . These generators are called Linear Feedback Shift Register generators (LFSR) or Tausworthe generators [4].

## 2 Floating Point Number Output

We consider a LCG like in Ex. 3 having maximal period length. The state space  $S = [1, m-1]$  then has cardinality  $\#S = m-1$  and all states occur with the same frequency  $\frac{1}{m-1}$ . We use the output function  $o(z) := z/m$ .

## 2.1 Moments of $U$

**Lemma 4.** For all LCGs with full period length  $\tilde{n} = m - 1$  we get:

$$E(U) = \frac{1}{2}$$

$$V(U) = \frac{1}{12} - \frac{1}{6m}$$

*Proof.* As the expected value  $E(U)$  and the variation  $V(U)$  are independent under permutations we simply compute:

$$E(U) = \sum_{n=1}^{m-1} \frac{n}{m} \cdot P(S = n) = \sum_{n=1}^{m-1} \frac{n}{m} \cdot \frac{1}{m-1}$$

$$= \frac{(m-1) \cdot m}{2} \cdot \frac{1}{m \cdot (m-1)} = \frac{1}{2}$$

$$V(U) = E(U^2) - E^2(U) = \frac{m - \frac{1}{2}}{3m} - \frac{1}{4} = \frac{1}{12} - \frac{1}{6m} .$$

We could therefore expect that these generators pass many tests of equidistribution. This is only nearly true as the following examples show.

## 2.2 Examples and Indications

We consider three basic examples from physics and electrical engineering.

**Example 5.** Consider a radio signal  $r(t) = \sin(2\pi \times f \times t)$  with frequency  $f$ . We want to determine the energy  $E = \int_0^1 \sin^2(2\pi \times f \times t) dt$  by Monte-Carlo integration and use the generator of Ex. 3 to produce  $n$  random numbers  $u_i \in (0,1)$  to simulate time instances  $t_i := 2\pi \times f \times u_i$ . With  $S := \sum_{i=1}^n \sin^2(t_i)$  one expects  $E \approx S/n$ . For a medium frequency signal (MF)  $f = 1048576 \text{ Hz} \approx 1 \text{ MHz}$  and  $n := 2^{20}$  numbers starting from seed 1 we get  $S/n = 0.50015$ . This is close to the true result  $E = 0.5$ . But for a micro wave signal (UHF) with frequency  $f = 1073741824 \text{ Hz} \approx 1 \text{ GHz}$  we get  $S/n = 0.0$ .

**Example 6.** We consider the signal  $r(t) = e^{-d \times t} \times \sin(2\pi \times f \times t)$  with damping  $d$  and calculate the signal energy  $E = \int_0^1 (e^{-d \times t} \times \sin(2\pi \times f \times t))^2 dt$  by Monte-Carlo integration. With the generator from Ex. 3,  $f = 1073741824 \text{ Hz}$ ,  $d = 1000$  and  $n := 2^{20}$  random numbers we again get  $S/n \approx 0.0$ .

**Example 7.** We consider the impulse  $r(t) = \exp(-d \times t^2) \times \sin(2\pi \times f \times t)$  with constant  $d$  and again determine the signal energy  $E = \int_0^1 r(t)^2 dt$  by Monte-Carlo integration. With the generator from Ex. 3,  $f = 1073741824 \text{ Hz}$ ,  $d = 1000$  or  $d = 100000$  and  $n := 2^{20}$  numbers we once more get  $S/n \approx 0.0$ .

### 2.3 Discrete Random Numbers

The output on a digital computer is a subset of finitely many discrete numbers  $o(S) \subseteq \mathbf{F} \subseteq \mathbf{R}$ . A finite set has measure 0 within the real numbers  $\mu(\mathbf{F})=0$  and we can immediately construct tests that all possible generators fail.

**Example 8.** Setting  $A := \mathbf{F} \cap (0,1)$  and  $B := (0,1) - \mathbf{F}$ , we get from the i.i.d. hypothesis  $F_0 = U(0,1)$  that  $P(A)=0$  and  $P(B)=1$ . As  $u_i \in A$  for all  $i=1,2,\dots$  all possible sequences in  $\mathbf{F}$  fail this test.

Therefore the random numbers generated can not to be i.i.d. real numbers. They have to be distributed according to a discrete distribution on a finite subset of computer numbers. We consider floating point numbers here (see [10]).

**Definition 9.** A floating point number  $v \in \mathbf{F}(b, p, e_{min}, e_{max})$  with basis  $b$ ,  $p$  significant digits, mantissa  $c$  and exponent  $e \in \{e_{min}, \dots, e_{max}\}$  has the value

$$v := (-1)^s \cdot c \cdot b^{-p} \cdot b^e \in \mathbf{Q}.$$

A *normal* floating point number  $v \neq 0$  has leading digit  $c_1 \neq 0$  and *denormal* floating point numbers have  $c_1 = 0$  and  $e = e_{min}$ .

**Example 10.** The IEEE *single* floating point format is  $\mathbf{F}(2, 24, -125, 128)$ .

## 3 Floating Point Number Tests

The usual goodness of fit tests like the  $\chi^2$ -square test or Kolmogorov-Smirnov test are not very sensitive to the kind of numbers used. These tests classify the numbers in relatively few equal length intervals, i.e. they are mainly determined by the first significant bits.

Floating point numbers are designed to minimize the relative error of computations. Tests should reflect this property. Better tests will use relative interval lengths. We propose tests of equidistribution for random floating point numbers based on intervals of the form  $[b^{p-1} \cdot b^e, (b^p - 1) \cdot b^e]$ .

### 3.1 Lower Bits Test

Testing the last significant bits of random numbers requires some care (see [1]). The states of LCG16807 in Ex. 3 can be called super uniform and even the lower bits are uniformly distributed. Taking  $n$  consecutive numbers and testing the lower bits with a  $\chi^2$ -test gives the probabilities  $p$  in Tables 1 and 2.

probability	occurrences	probability	occurrences
$p < 0.01$	0	$0.05 \leq p < 0.10$	8
$0.01 \leq p < 0.05$	4	$0.10 \leq p$	68

Table 1. 80  $\chi^2$ -tests: 3 lower bits,  $2^{17}$  numbers



probability	occurrences	probability	occurrences
$p < 0.01$	0	$p < 0.01$	0
$0.01 \leq p < 0.10$	2	$0.01 \leq p < 0.10$	1
$0.10 \leq p$	8	$0.10 \leq p$	9

Table 2. 10  $\chi^2$ -tests: 3, 6 lower bits,  $2^{20}$  numbers

### 3.2 Mantissa Test

The *mantissa test* classes all significands  $c$  according to their lower bits.

We take the random number generator LCG16807 of Ex. 3, output in  $F(2, 24, -125, 128) \cap [0, 1]$  and the usual output function  $o(z) := z/m$ . Again we test  $n$  consecutive random numbers with a  $\chi^2$ -test taking the last bits of the mantissa  $c$ . The results are shown in Tables 3 and 4.

probability	occurrences	probability	occurrences
$p < 0.01$	0	$0.05 \leq p < 0.10$	9
$0.01 \leq p < 0.05$	1	$0.10 \leq p$	0

Table 3. 10  $\chi^2$ -tests: 3 lower bits,  $n = 2^{17}$  numbers

probability	occurrences	probability	occurrences
$p < 0.001$	7	$p < 0.001$	2
$0.001 \leq p < 0.01$	2	$0.001 \leq p < 0.01$	2
$0.01 \leq p < 0.10$	1	$0.01 \leq p < 0.10$	4
$0.10 \leq p$	0	$0.10 \leq p$	1

Table 4. 10  $\chi^2$ -tests: 3, 6 lower bits,  $n = 2^{20}$  numbers

Clearly the generator LCG16807 with this output function fails our test. The reason for this is, that not all the mantissa for small floating point numbers are generated.

This can be seen even clearer if one uses  $m = 2^{31} - 1 \approx 2^{31}$  and the output function  $o(z) := \text{round}(z) \cdot 2^{-31}$ , rounding to  $p$  significant digits and right shift  $2^{-31}$ . Due to normalization all floating point mantissa for values smaller than  $1/256$  have last bit 0.

## 4 Floating Point Number Generation

As the usual output functions lead to generators failing floating number tests we have to construct new generators passing these tests.

T. Morgenstern

A new generator FloatLCG2 based on LGC16807 generates floating point numbers  $(-1)^s \cdot c \cdot b^{-p} \cdot b^e$ , by generating the exponent  $e$  and the mantissa  $c$  with  $p$  significant digits independently in two steps. It also generates denormal numbers.

**Step 1 (exponent generation):**

```
denormal:=true;
e:=0;
while (e ≥ emin and denormal) do
  s:=16807*s mod 2147483647;
  i:= 30; while (s < 2i and i > 0) do i:= i-1 end do;
  e:=e+i-30;
  if i > 0 then denormal:=false end if;
end do;
if e < emin then e:=emin; denormal:=true end if;
```

**Step 2 (mantissa generation, rounding):**

```
s:=16807*s mod 2147483647;
if (not denormal) then c:=231+s else c:=s end if;
c:=c*2p-32;
c:=trunc(c);
u:=c*2e-p;
```

## 5 Empirical Tests

This new binary floating point number generator FloatLCG2 has to pass the usual statistical tests.

### 5.1 $\chi^2$ -Test

A description of this test can be found in [1]. We take  $n = 131072$  numbers for a test and use 16384 equal length intervals, i.e. the expected cell frequency is 8. 30 test replications give 4 times  $0.01 \leq p < 0.05$ , 3 times  $0.05 \leq p \leq 0.10$  and 23 times  $0.10 < p$  for the probability  $p$  of the sequence, i.e. non-suspicious values. The generator passes this test.

### 5.2 Kolmogorov-Smirnov Test

A description of this test can be found in [1]. Again we take  $n = 131072$  numbers per test. The results of 100 tests can be found in Table 5.

probability	occurrences	probability	occurrences
$p^- < 0.01$	1	$p^+ < 0.01$	1
$0.01 \leq p^- < 0.05$	2	$0.01 \leq p^+ < 0.05$	5
$0.05 \leq p^- < 0.10$	6	$0.05 \leq p^+ < 0.10$	5
$0.10 \leq p^-$	81	$0.10 \leq p^+$	79

Table 5. KS tests with  $n = 2^{17}$  numbers

If we take  $n = 1048576$  numbers and 20 tests we only get for seed 698 499 821 the low value  $p^- = 0.038$ , all other values being even less suspicious. This test is passed.

### 5.3 Serial Correlation Test

A description of this test can be found in [1]. We take  $n = 131072$  numbers in one test and lags from 1 to 100. Starting from seed 1 the correlations found lie between  $-0.00573$  and  $0.00573$ . The  $\chi^2$ -test as well finds no indication against independency of the shifted series. The generator passes these tests.

### 5.4 Mantissa Test

We take 20 tests with  $n = 1048576$  numbers starting from seed 1 and test the lower 1, 3 and 6 bits of the mantissa with the  $\chi^2$ -test. The lowest probability found once for the last 2 bits test is  $p = 0.092$ . Testing the last 3 bits one once finds  $p = 0.098$ . All the test values are non-suspicious. Our floating point generator FloatLCG2 also passes this new test.

## 6 Results

**Example 11.** We continue Ex. 5 using our new generator FloatLCG2 to compute  $u_i \in (0,1)$ . For frequency  $f = 1073741824$  Hz,  $n := 2^{20}$  and seed 1 we still get the only slightly improved value  $S/n = 0.001990$  for the Energy.

**Example 12.** Continuing Ex. 6, the damped signal example, with FloatLCG2,  $n = 2^{20}$  and seed 1 we get the fairly good value  $S/n \approx 0.000253$ . 7 further repetitions of this simulation give values between 0.000240 and 0.000262, in average 0.0002490 (i.e. for  $n = 2^{23}$ ), near to the true value 0.000249999999999999945.

**Example 12.** We continue Ex. 7, the ms-impulse example, with the new generator FloatLCG2,  $n = 2^{20}$  and seed 1. For  $d = 1000$  we get the improved value  $S/n \approx 0.00197$ , closer to the approximate true value 0.009908. With  $d = 100000$  we get  $S/n \approx 0.000979$ . In 7 further repetitions with  $d = 100000$  we get values between 0.000933 and 0.001043, in average 0.000976 (i.e. for  $n = 2^{23}$ ), close enough to the approximate true value 0.0009908.

The three examples show that some basic simulation tasks in physics or engineering fail, when classical pseudorandom number generators are used. In applications the systems are

T. Morgenstern

much more involved and wrong simulation results might not be recognised, even if the results are far from the true values. The new pseudorandom number generator FloatLCG2 for the generation of uniform random binary floating point numbers is better and solves a wider class of problems. But further work is required to solve problems for which this generator is still not appropriate [11].

## References

- [1] KNUTH, D. E., *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, 3rd ed. Addison-Wesley, Reading, MA. (1998).
- [2] L'ECUYER, P., Uniform random number generation, *Annals of Operations Research* (1994) 77–120.
- [3] L'ECUYER, P. Random Number Generation, *Handbook of Computational Statistics*, Gentle, J. E., Haerdle, W., and Mori, Y., Eds., Springer-Verlag (2004) 35–70.
- [4] L'ECUYER, P. Random Number Generation, *Handbook on Simulation*, Jerry Banks Ed., Wiley (1998) 93–137.
- [5] L'ECUYER, P., Software for Uniform Random Number Generation: Distinguishing the Good and the Bad, *Proceedings of the 2001 Winter Simulation Conference*, IEEE Press (2001) 95–105.
- [6] MAPLE 10, Maplesoft a division of Waterloo Maple Inc., [www.maplesoft.com](http://www.maplesoft.com) (2005).
- [7] LEWIS, P.A.S., GOODMAN, A.S., MILLER, J.M., A pseudo-random number generator for the system/360, *IBM System's Journal* (1969) 136–143.
- [8] LEARMONTH, G.P., LEWIS, P.A.S., Naval Postgraduate School Random Number Generator Package LLRANDOM, Monterey, California (1973).
- [9] DUDEWICZ, E.J., RALLEY, T.G., The handbook of random number generation and testing with TESTRAND computer code, American Sciences Press, Columbus, Ohio, (1981).
- [10] GOLDBERG, D., What Every Computer Scientist Should Know About Floating-Point Arithmetic, *ACM Computing Surveys*, 23, (1991) 5– 48.
- [11] MORGENSTERN, T., Uniform Random Rational Number Generation, *Operations Research Conference 2006 in Karlsruhe*, Gesellschaft für Operations Research GOR (submitted).

# Augmented Learning in der optischen Nachrichtentechnik – Konzeption einer interaktiven Lernsoftware zur Unterstützung der selbständigen Laborarbeit

Dipl.-Ing.(FH) Jens-Uwe Just, Phone: ++49.3943-659-340, Fax: ++49.3943-659-399,  
E-Mail: jjust@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

Dipl.-Wilnf.(FH) Christian Reinboth, Phone: ++49.3943-659-340,  
E-Mail: creinboth@hs-harz.de, Hochschule Harz

Dipl.-Ing. Matthias Haupt, Phone: ++49.3943-659-368,  
E-Mail: mhaupt@hs-harz.de, Hochschule Harz

Prof. Dr.-Ing. habil. Ulrich H.P. Fischer-Hirchert, Phone: ++49.03943-659-105,  
E-Mail: ufischerhirchert@hs-harz.de, Hochschule Harz

**Abstract:** Dieses Paper beschreibt die Konzeption einer Lernsoftware für den Einsatz in Laborveranstaltungen der optischen Nachrichtentechnik. Es wird dargestellt, auf welchen Grundlagen die Entscheidungen bezüglich der technischen und inhaltlichen Konzeption der Software getroffen wurden, und wie das didaktische Konzept in den drei Merkmalsdimensionen von Instruktionssystemen seine Ausprägung findet.

**Keywords:** Rechnergestützte Laborarbeit, Lernsoftware, Augmented Learning

## 1 Einführung

Aktuelle Kommunikationssysteme basieren meist auf fortgeschrittenen opto-elektronischen Verfahren, wobei sich das Wellenlängenmultiplex (WDM<sup>1</sup>) als besonders aussichtsreich erwiesen hat [Fi02]. Als Lichtwellenleiter lassen sich neben Glasfasern auch Polymerfasern (POF<sup>2</sup>) einsetzen, die als künftiger Quasi-Standard im Kurzstreckenbereich gelten [We02]. Die Vermittlung von Kenntnissen über POF und WDM wird eine der künftigen Aufgaben von technisch orientierten Bildungseinrichtungen sein.

Die Vermittlung theoretischer und praktischer Kenntnisse erfolgt dabei meist getrennt. Während die Theorie in Vorlesungen vermittelt wird, erfolgt die praktische Ausbildung in Laborveranstaltungen mit Lehrsystemen, welche die praktische Umsetzung theoretischer Kenntnisse ermöglichen. Die Lehrenden sind dabei mit dem Problem konfrontiert, sich an die unterschiedlichen Fortschritte der Laboranten anpassen zu müssen. Häufig ist es während der Versuche erforderlich, theoretische Inhalte aufzugreifen oder individuell Hilfe zu leisten, was nicht dem Wunschbild einer hohen Autonomie der Lernenden entspricht. Statt dessen sollten die Lernenden parallel zum Versuchsablauf theoretische Inhalte nachschlagen oder bei Problemen autonom zu einer Lösung gelangen können. Der Einsatz von Lernsoftware kann dazu beitragen, die Laborarbeit näher an dieses Ideal heranzuführen.

In diesem Paper wird die Konzeption einer solchen Lernsoftware beschrieben, die parallel zu einem WDM-POF-Lehrsystem an der HS Harz entwickelt wird. Das System (Abb.1) besteht aus zwei Sendern und Empfängern, mit denen sich die Übertragung von analogen FBAS-Signalen mit 10MHz über 20m POF realisieren lässt. Der WDM-Effekt wird dabei durch Y-Koppler der Firma DieMount erzielt, wobei der Ersatz durch Eigenentwicklungen geplant ist,

---

<sup>1</sup> WDM, Wavelength Division Multiplex

<sup>2</sup> POF, Polymer Optical Fibre

die sich noch in der Patentierung befinden. Es ermöglicht eine Vielzahl von Versuchen zur Nachrichtentechnik, darunter Messung von Bandbreite und S-Parameter und Analyse des Einflusses von Faserbrüchen und Stoßstellen.

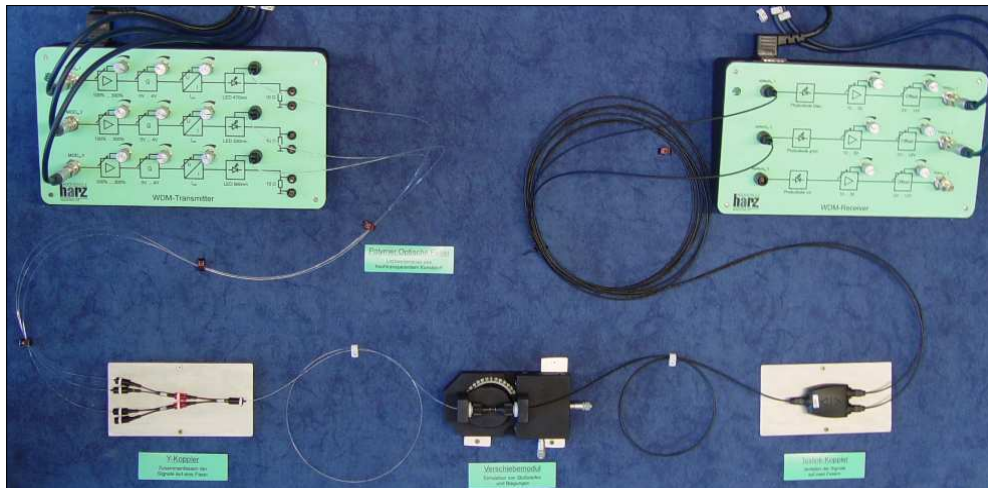


Abb.1: Prototyp des optischen Lehrsystems, das durch die Lernsoftware ergänzt werden soll

## 2 Technische Umsetzung

Für Lernsoftware lassen sich zwei wesentliche technische Anforderungen feststellen: Plattform-Unabhängigkeit und Multimedialität. Es muss gewährleistet sein, dass die Software unabhängig von den verfügbaren Endgeräten lauffähig ist. Daher lassen sich ausschließlich nicht-proprietäre Technologien wie Java oder HTML einsetzen. Bei Verwendung von HTML ist zu beachten, dass sich die Endgeräte durch unterschiedliche Browsertypen und Bildauflösungen unterscheiden können. Die Software muss daher auf verschiedenen Konfigurationen getestet und ggf. angepasst werden. Die Entwicklung speziell für eine Kombination aus Betriebssystem, Browsertyp und Bildauflösung ist nicht ratsam, da sie die Nutzer dazu zwingt, ihre Technik den Herstellervorgaben anzupassen.

Auch die Multimedialität ist auf der technischen Ebene umzusetzen. Anwender heutiger Software erwarten multimedial umgesetzte Inhalte, zudem hilft deren Einbindung die Lerneinheiten ansprechender zu gestalten und die Aufmerksamkeit der Lernenden zu aktivieren [Te00]. So können kurze Videosequenzen aus der Laborpraxis zur Veranschaulichung theoretischer Sachverhalte oder zur Darstellung der Versuchsabläufe dienen. Beim Entwurf eines HTML-basierten Multimediakonzepts ist darauf zu achten, dass keine unüblichen Browser-Plugins verwendet werden, die aus lizenzrechtlichen Gründen nicht mit der Software weitergegeben werden können. Das Softwarekonzept sieht daher den ausschließlichen Einsatz von Techniken vor, die keine Plugins voraussetzen (z.B. animierte GIFs) oder Plugins benötigen, mit deren Vorhandensein zu rechnen ist (z.B. Flash).

Analog zur Umsetzung von Online-Befragungen stellt sich bei Lernsoftware auch die Frage, ob zusammenhängende Inhalte auf einer Seite dargestellt werden sollten oder ob die Trennung in Informationseinheiten vorzuziehen ist, die sich auf einer Bildschirmseite unterbringen lassen. Wird die gesamte Lerneinheit auf einer Seite präsentiert, so ist der Kontext für den Befragten unmittelbar ersichtlich. Das sich negativ auf die Motivation auswirkende Gefühl der „Endlosigkeit“ kann nicht aufkommen, da der Bearbeitungsaufwand eingeschätzt werden kann. Die Programmierung eines solchen Systems ist weniger

aufwendig als bei Single-Screen-Systemen, so dass sich die Inhalte schneller umsetzen lassen. Es ist jedoch bekannt, dass diese Form der Darstellung zum schnellen und unkonzentrierten „Durchscrollen“ der Inhalte verführt [Te00].

Werden dagegen kleine Informationseinheiten gebildet, so muss der Lernende nicht scrollen, die Navigation ist übersichtlicher und komfortabler [Te00]. Der Sprung von Seite zu Seite ermöglicht zudem den Einsatz von Kontrollfragen, die unmittelbar überprüft werden können, so dass der Lernende ein direktes Feedback erhält und seinen Lernerfolg realistisch beurteilen kann. Der Grad der Interaktivität sowie die Kontrollmöglichkeiten sind bei der Single Screen Variante deutlich größer, weshalb der höhere Programmieraufwand akzeptabel ist. Um das Gefühl der Unüberblickbarkeit nicht aufkommen zu lassen, ist geplant, eine Fortschrittsanzeige einzubinden, außerdem wird die durchschnittliche Bearbeitungszeit jedes Moduls bei Aufruf eingeblendet werden.

Die Navigation innerhalb der Inhalte sollte dem Lernenden eine größtmögliche Übersicht bieten und sowohl das einfache Springen zwischen den Inhalten gestatten, als auch den didaktisch gewünschten Pfad hervorheben. Analog zu Online-Befragungen wird ein Pretest durchgeführt werden, um die Annahme der Navigation durch die Anwender zu überprüfen. Die grundsätzliche Methodik dieses Tests kann aus dem Befragungsbereich übernommen werden und findet sich z.B. in [Po98].

### **3 Didaktische Umsetzung**

Da die Software den gesamten Lernprozess begleiten soll, ist festzustellen, welche Phasen sich nach methodischen Gesichtspunkten unterscheiden lassen. Im vorliegenden Fall sind diese aus den Erfahrungen mit Laborveranstaltungen bereits bekannt: Wiederholung theoretischer Inhalte, Anleitung und Begleitung der Laborversuche, Erfassung von Messdaten und Erstellung der Laborprotokolle.

Die Wiederholung der theoretischen Inhalte zieht sich nach einer knappen Einführung durch Versuche und Auswertungen. HTML gestattet die Umsetzung des Hypertext-Prinzips in der Software, da aus den Versuchsanleitungen jederzeit auf theoretische Inhalte zurückgesprungen werden kann. Ein ständig über die Navigationsleiste verfügbares Glossar mit Suchfunktion wird den permanenten Informationszugriff zusätzlich erleichtern. Der Synergieeffekt, der sich aus der Kombination von selbst erarbeiteten Ergebnissen und theoretischer Wiederholung ergibt, kann so optimal ausgenutzt werden. Da das Single Screen-Modell die Einbindung von Zwischenfragen ermöglicht, kann das Verständnis der Lernenden zusätzlich überprüft werden. Dem Lehrenden wird es möglich sein festzulegen, ob die Beantwortung der Fragen mandatorisch oder freiwillig sein soll.

Die Begleitung der Laborversuche ist das Herzstück der Software. Parallel zur Durchführung soll der Versuch am Bildschirm nachvollzogen werden, wobei durchschnittliche Messdaten die permanente Selbstkontrolle ermöglichen. Die Lernenden werden so in die Lage versetzt, Abweichungen vom idealen Versuchsablauf frühzeitig zu erkennen. Im Falle deutlicher Abweichungen werden die Lernenden zudem die Möglichkeit haben, über ein Fragesystem gängige Ursachen für die aufgetretenen Probleme zu ermitteln und sich Instruktionen zu deren Behebung anzeigen zu lassen.

Die Erfassung der Messdaten und Erstellung der Laborprotokolle wird ebenfalls in der Software umgesetzt. Über ein Formularfeld können Messdaten und Textantworten auf Verständnisfragen eingegeben werden. Dies bedeutet eine Arbeitserleichterung für die Lehrenden, die keine handschriftlichen Protokolle mehr auswerten müssen, sondern die auf einem Server gesammelten Protokolldateien aufrufen und auf deren Basis die Testate erteilen können.

Didaktische Systeme sind durch die drei Merkmale Interaktivität, Adaptivität und Kontrollinstanz beschreibbar [DE01]. Interaktivität ist nach [Ke98] eine technische Eigenschaft: In einem interaktiven Medium besteht ungehinderter Informationszugriff. Dies gestattet die aktive inhaltliche Verarbeitung durch den Anwender, der durch seine Aktionen Auswahl und Abfolge der Inhalte kontrollieren kann [Ja00]. Die Interaktivität der Lernsoftware wird durch die leicht erlernbare Navigation gegeben sein, welche den unmittelbaren und uneingeschränkten Zugriff auf alle Informationen gestattet.

Adaptivität bezeichnet das Maß, in dem Lernende eine Lernumgebung an ihre Bedürfnisse anpassen können [DE01]. Die Software wird es Lernenden gestatten, in ihrer persönlichen Arbeitsgeschwindigkeit zu navigieren, Passagen beliebig zu wiederholen, sich durch Kontrollfragen zu prüfen und während der Versuche die Instruktionsgeschwindigkeit an den tatsächlichen Ablauf anzupassen.

Die Kontrollinstanz bezieht sich auf die Möglichkeit der Lernenden, Einfluss auf den Lernvorgang auszuüben [DE01]. Für computergestützte Lernumgebungen gilt nach [Ne00], dass die Kontrollinstanz mit dem Grad der nichtlinearen Navigation ansteigt, d.h. je stärker sich der Nutzer frei durch die Inhalte bewegen kann, desto deutlicher ist diese Dimension ausgeprägt. Aufgrund der Hypertext-Funktionalität der Lernsoftware ist eine nahezu uneingeschränkte Navigation durch die Inhalte möglich. Restriktionen wird es lediglich bei der Erstellung der Protokolle geben, die als Prüfungsleistung gewertet werden, und daher zeitlichen und inhaltlichen Vorgaben unterliegen.

#### **4 Fazit und Ausblick**

Die Softwareentwicklung wird nach aktuellem Stand bis November 2006 abgeschlossen sein. Das optische Lehrsystem existiert bereits als Prototyp, eine Nullserie befindet sich zur Zeit in Produktion. Mit der Patenterteilung für die selbst entwickelten Funktionsbauteile ist bis Jahresende zu rechnen.

Da das hier dargestellte Konzept der softwaregestützten Laborarbeit für den Bereich der optischen Nachrichtentechnik als neu zu betrachten ist, wird in enger Zusammenarbeit mit den ersten Abnehmern ein Quality Feedback System eingerichtet, welches sowohl Lehrenden als auch Lernenden die Möglichkeit bieten wird, Erfahrungen, Lob und Kritik zu kommunizieren. Neben subjektiven Eindrücken werden dabei auch objektive Daten wie Noten und Lehrevaluations-Ergebnisse erhoben.

Die kontinuierliche Weiterentwicklung von Lehrsystem und Lernsoftware sowie die Dokumentation dieses Prozesses, wird eine umfassende Darstellung der Anforderungen und Probleme beim Einsatz interaktiver Lernsoftware im Laborbereich ermöglichen. Es besteht die Hoffnung, aus den Erfahrungsberichten und der Dokumentation des Entwicklungsprozesses einen umfangreichen Anforderungskatalog für eine solche Software erstellen und veröffentlichen zu können.



## Quellenangaben

- [DE01] Dichanz, H; Ernst, A: E-Learning. Begriffliche, psychologische und didaktische Überlegungen zum „electronic learning“, [www.medienpaed.com](http://www.medienpaed.com)
- [Fi02] Fischer, U.H.P.: Optoelectronic Packaging, VDE-Verlag, 2002
- [Ja00] Jarz, E.: Entwicklung multimedialer Systeme. Planung von Lehr- und Masseninformati<sup>o</sup>nssystemen. DeutscherUniversitätsVerlag, 2000
- [Ke98] Kerres, M.: Multimediale und telemediale Lernumgebungen. Konzeption und Entwicklung. München, Wien, Oldenbourg-Verlag, 1998
- [Ne00] Niegermann, H.: Konzeption, Entwicklung und Einsatz digitaler Lernumgebungen. In: Beiträge aus dem wissenschaftlichen Leben: Medien, TU Ilmenau, Ilmenau, 2000, S. 16
- [Po98] Porst, R.: Im Vorfeld der Befragung: Planung, Fragebogenentwicklung, Pretesting. In: ZUMA-Arbeitsbericht 98/02, Mannheim, ZUMA, 1998
- [Te00] Theobald, A.: Das WWW als Befragungsinstrument. Wiesbaden, Gabler-Verlag, 2000
- [We02] Weber, N.: Low cost optical transmission solutions for short distances, <http://www.iis.fraunhofer.de/ec/oc/index.html>



# **Eingeladener Beitrag II**



# Wissen – Ziele – Aktion: Eine unabdingbare Folge – auch in der Automation

Hartwig Steusloff

Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB), Karlsruhe

## Kurzfassung:

Die Einsicht, dass Wissen allein ohne die Berücksichtigung von wohl definierten Zielen nicht zu erfolgreichen Aktionen führen kann, wirft die Frage nach der Art solcher Zieldefinitionen – speziell auch für die Automation – auf. Zielssysteme, wie sie Gegenstand von Ethiken sind, können hier weiterhelfen, insbesondere auch für die unabdingbare Einbeziehung des Menschen in Systeme der Automation.

## 1 Die Kette von den Daten zur Aktion

Aktionen basieren auf Kompetenzen, die durch Wissen abgestützt sind. Wissen wiederum wird abgeleitet aus Informationen und Daten (Bild 1). Diese Zusammenhänge gelten für alle Arten von Aktionen unter der Voraussetzung, dass verfügbares Wissen und verfügbare Kompetenzen für die notwendigen oder erwünschten Aktionen passend und ausreichend sind. Ist dies schon hinsichtlich des geeigneten Wissens nicht immer eindeutig feststellbar, so dürfte die Eignung der in die Kompetenz eingehenden Erfahrungen und Ziele bezüglich anstehender Aktionen oft noch weniger klar sein. In diesem Beitrag soll der Versuch gemacht werden, die systematische Definition der notwendigen oder zumindest geeigneten Kompetenzen, insbesondere der damit verbundenen Zielsetzungen, für die Automation technischer Systeme zu untersuchen.

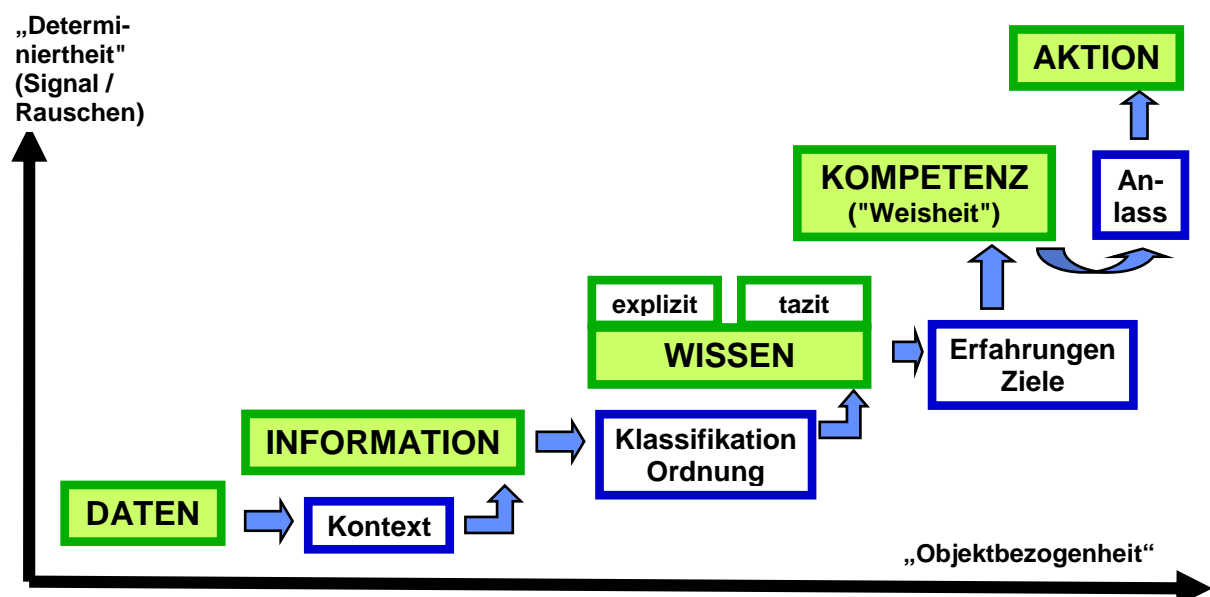


Bild 1: Von Daten zur Aktion [1]

## 2 Automation

Die Grundstruktur der Automation wurde durch den Beirat der VDI/VDE-Gesellschaft für Mess- und Automatisierungstechnik (GMA) definiert und ist in Bild 2 dargestellt. Wesentlich ist die Unterteilung in die geschlossene funktionale Kette vom (technischen) Prozess über die Informationsgewinnung, die Informationsverarbeitung und die Informationseinwirkung zurück zum Prozess ("Automatisierung") sowie – als unabdingbare Erweiterung – die Einwirkung des Menschen auf den Prozess über die Informationsverarbeitung zum Zwecke der Mensch-Prozess-Kommunikation ("Automation"). Dieser Kern der Automation ist umgeben von den einzuhaltenden Vorgaben und Restriktionen und den erforderlichen Infrastruktursystemen der Informationslogistik sowie der Entwurfsunterstützung; Automation basiert schließlich auf Grundlagen und Methoden, die aus der Forschung und Entwicklung kommen.

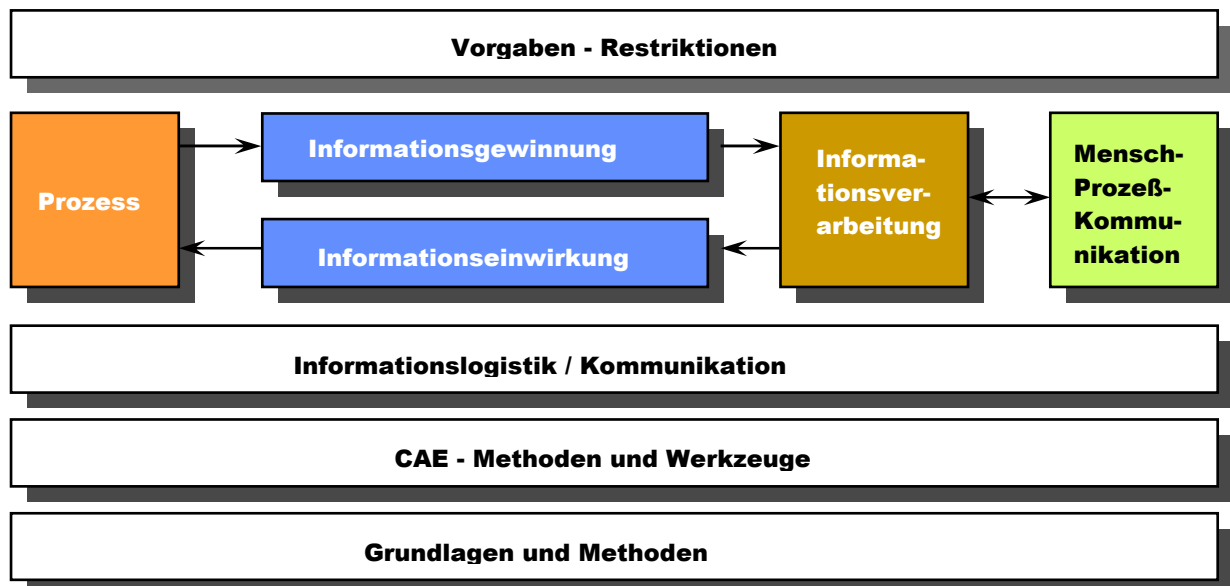


Bild 2: Grundstruktur der Automation

Diese Struktur umfasst zahlreiche Aktionen der Forschung und Entwicklung, des Entwurfes und der Planung sowie der Implementation und des Betriebes eines Automationssystems. Insbesondere bezieht sie aber auch explizit den Menschen als Teil des Systems ein. Die Notwendigkeit, den Menschen durch eine geeignete Mensch-Prozess-Kommunikation zum Teil eines Automationssystems zu machen, ergibt sich aus der Tatsache, dass die Automatisierung nur diejenigen Prozesszustände beherrschen kann, die explizit und algorithmisch beschreibbar sind. Alle weiteren Prozesszustände bleiben dem Eingreifen des Menschen vorbehalten, der auf der Basis seiner Erfahrungen und der für den Prozessablauf gesetzten Ziele handeln muss. Die Kompetenz des Menschen ist damit unverzichtbarer Bestandteil eines jeden Automationssystems.

Daraus ergibt sich die Frage, wie diese Kompetenz des Menschen entwickelt werden kann und welche Ziele dabei zu berücksichtigen sind. Während die Ziele der Automatisierung, wie

- verlässlicher Prozessbetrieb,

H. Steusloff

- sicherer Prozessbetrieb oder
- effizienter Prozessbetrieb

durch entsprechende technische Maßnahmen sichergestellt werden können, sind übergeordnete Ziele, wie

- hohe Produktivität oder
- flexible Prozessführung

vorwiegend menschlichen Aktionen zugeordnet. Darüber hinaus ist der Mensch für die

- Beherrschung *sämtlicher* Prozesszustände

verantwortlich, d.h. auch derjenigen Prozesszustände, die im Automatisierungssystem nicht behandelt oder abgefangen werden und werden können. Die Kompetenz des menschlichen Prozessoperators muss also sowohl die "Kompetenzen" des (vom menschlichen Systemingenieur geplanten und implementierten) Automatisierungssystems wie auch darüber hinaus gehende Kompetenzen zur Beherrschung unvorhergesehener oder unvorhersehbarer Prozesszustände umfassen.

Hier stellt sich die Frage nach der Definition und Ausbildung der notwendigen Kompetenzen des Menschen als Teil eines Automationssystems, wobei wir hier bewusst die in gleicher Weise zu stellende Frage nach den notwendigen Kompetenzen des menschlichen Systemplaners und –errichters nicht betrachten wollen. Auch sei im Folgenden voraus gesetzt, dass die heute vorliegenden Erkenntnisse der Ergonomie und Anthropotechnik [2] bei der Gestaltung der Interaktionsschnittstelle zwischen dem Menschen und der Informationsverarbeitung (siehe Bild 2) beachtet worden sind. Dennoch bleiben für die Kompetenz des für die Prozessführung verantwortlichen Operators eine Vielzahl von Fragen zu beantworten, auf die wir im Folgenden eingehen wollen.

Es mag reizvoll sein, die mit der Definition, der Errichtung und dem Einsatz von Kompetenzen verbundenen Ziele als eine Ethik aufzufassen, die ja ebenfalls Ziele für Aktionen innerhalb menschlicher Gesellschaften definiert. Dazu wollen wir zunächst der Frage nachgehen, was eine solche Ethik bedeutet und leisten kann.

### **3 Ethik: Bedeutung und Konzepte**

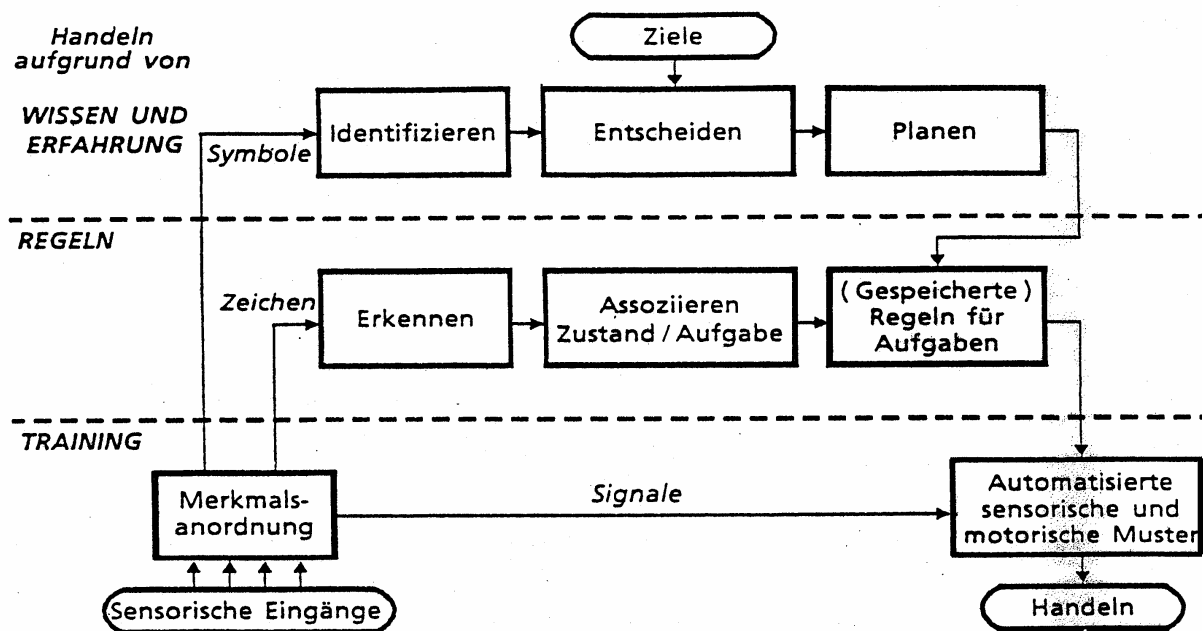
Das Wort "Ethik" ist abgeleitet aus dem griechischen Wort "Ethos(ἦθος)" und bedeutet "Wohnsitz, gewohnter Lebensort, Brauch, Sitte" wie auch "kultivierte Lebensform". Damit ist das Grundanliegen einer jeden Ethik, Leitlinien bzw. Orientierung für das Gelingen eines "guten" Lebens zu geben. Durch den Hinweis auf eine kultivierte Lebensform und ein gutes Leben ist damit gleichzeitig eine Wertung dieses Grundanliegens gegeben: Die Ethik soll zum Gelingen des Lebens beitragen! Im Gegensatz zur Moral, die sich faktisch über Emotionen begründet sowie kultur- und gesellschaftsabhängig ist, will die Ethik allgemein anwendbare Leitlinien für das Gelingen des Lebens systematisch entwickeln.

Unter den vielfältigen Arten von Ethik ist für die Betrachtung technischer Systeme insbesondere die so genannte Verantwortungsethik wichtig. Diese stellt die Handlungen und ihre absehbaren Folgen in den Mittelpunkt und postuliert Aktionen, deren Folgen möglichst viele positive Wirkungen sind, ohne das Auftreten von negativen Wirkungen zu ignorieren. Damit gehört die Verantwortungsethik zu den wertenden Arten der Ethik.

Der Wunsch alles Lebendigen, sich zu entfalten, erzeugt einerseits Rivalität bei der Beanspruchung von Ressourcen, begründet andererseits aber auch Kooperation bei der Bewältigung von Problemen und bei der Beschaffung von Ressourcen. Die Ethik als Hilfe zum Gelingen eines guten Lebens muss die Bewältigung dieses Antagonismus unterstützen u.a. durch die Forderung nach Respekt und Wohlwollen speziell gegenüber Mitmenschen und anderen Lebewesen sowie die Abschätzung und Bewertung der Folgen jeglicher Handlung. Dies gilt für das persönliche Handeln ebenso wie für die Etablierung von Strukturen und Regelungsmechanismen für die Verteilung von Ressourcen, die Verteilung von Lasten, von Chancen und von Einschränkungen, von Macht und von Verantwortung. Damit sind wir bei den Überlegungen zur Definition und Anwendung ethischer Konzepte im Rahmen der Automation.

#### 4. Der Mensch in der Automation

Wie Bild 2 deutlich macht, ist der Mensch in einem System der Automation unverzichtbar für das Erreichen der gesetzten Ziele. Eine Formulierung solcher auf den Menschen bezogenen Ziele setzt voraus, die entsprechenden Eigenschaften des Menschen zu kennen und zu berücksichtigen. Es existieren zahlreiche Modelle für die Eigenschaften und das Verhalten des Menschen im Kontext eines (technischen) Systems, von denen hier nur zwei Modelle erwähnt werden sollen.



3-Schichtenmodell des menschlichen Handelns ( nach Goodstein )

Bild 3: Verhaltensmodell des Menschen nach Rasmussen/Goodstein [3]

Das Modell von Rasmussen/Goodstein beschreibt das Verhalten des Menschen in drei Ebenen, der untersten intuitiven Ebene, der zweiten regelgesteuerten Ebene und der obersten zielgesteuerten Ebene. Für unsere ethik-orientierte Betrachtung ist vor allem die ziel-



gesteuerte Ebene wichtig: Der Mensch handelt auf Grund von Wissen und Erfahrung unter Berücksichtigung von Zielen und kann dadurch auch unerwartete Systemzustände behandeln. Voraussetzung ist die Existenz solcher Ziele, die je nach Komplexität der zu behandelnden Situation ausreichend allgemein formuliert sein müssen und damit die Charakteristika einer Ethik aufweisen.

Als zweites Modell sei ein Belastungsmodell, das Zeitbudgetmodell, erwähnt, welches die Belastung des Menschen bei der Abarbeitung vielfältiger Anforderungen beschreibt. Dieses Warteschlangenmodell impliziert ebenfalls Entscheidungen bei der Abarbeitung des Anforderungsstromes und weist auf ein wichtiges Erfordernis bei der Einbindung des Menschen in den Prozessbetrieb hin: Die Belastung darf den Menschen weder über- noch unterfordern. Das verfügbare Zeitbudget wird verbraucht durch eine Anzahl unterschiedlicher Aktionen, die für typische Anforderungsprofile aus Büro- und Automationssystemen in Bild 5 dargestellt sind [4].

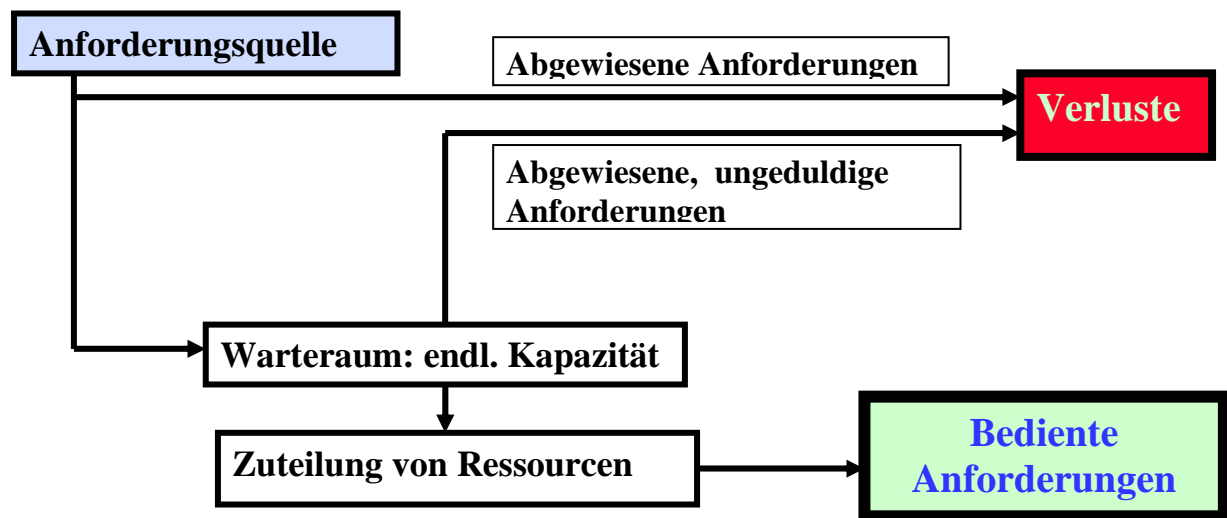


Bild 4: Zeitbudgetmodell

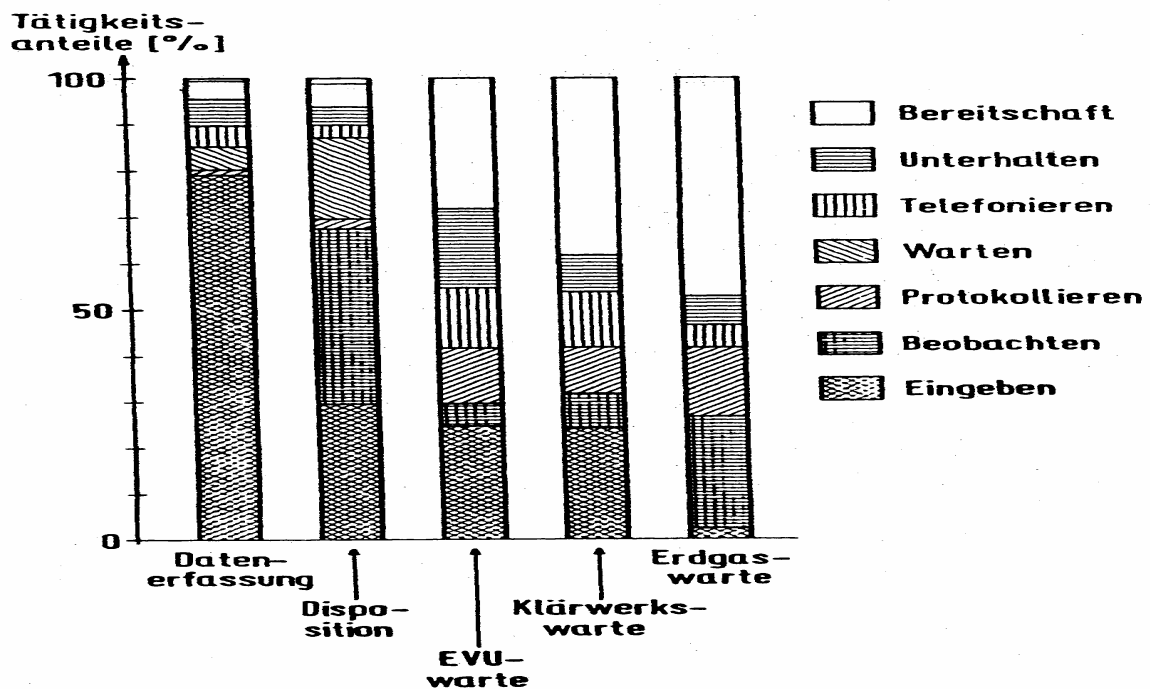


Bild 5: Aktionen und Zeitbudget für verschiedene Anforderungsprofile [4]

Bild 5 macht deutlich, dass sowohl monotone Aufgaben, wie z.B. das Eingeben von großen Datenmengen, als auch Unterforderungen durch sehr große Anteile von Bereitschaft auftreten.

## 5 Ansatz einer mensch-bezogenen Ethik für die Automation

Aus den vorstehenden Konzepten und Modellen für eine Automation mit Einbezug des Menschen lassen sich Ziele im Sinne einer Verantwortungsethik ableiten und begründen. Im Folgenden sei ein Versuch unternommen, eine solche Verantwortungsethik in Teilen zu skizzieren. Diese soll sich hier ausschließlich auf den in die Automation eingebundenen Menschen beziehen und Verantwortung für Gesellschaft und Politik, für die Umwelt, für die Schonung von Ressourcen u.v.a.m. nicht explizit einschließen. Dies bedeutet natürlich nicht, dass die folgenden Leitlinien keinen Einfluss auf die Gesellschaft, die Umwelt und die Ressourcenschonung hätten. Auch können die folgenden Überlegungen keinesfalls und auch nur annähernd Anspruch auf Vollständigkeit erheben.

Ein wichtiger Ausgangspunkt ist die Arbeitsethik. Der Schweizer Theologe und Sozialethiker Alfred de Quervain hat vier Punkte definiert, die für ein gutes Arbeitsleben gegeben sein müssen [5]:

1. Arbeit darf nicht fremdbestimmend sein im Sinne einer Vorherrschaft der technischen Umgebung des Menschen. Arbeit muss Gestaltungsspielräume geben und dazu dem Menschen Verantwortung übertragen und diese herausfordern.  
Für die Automation heisst dies, dass der Mensch nicht lediglich Bediener einer automatisierten Anlage sein soll sondern, im Sinne von Rasmussen und Bild 3, möglichst viele zielorientierte Gestaltungsmöglichkeiten haben muss.

2. Arbeit muss Lebensgestaltung und Selbstverwirklichung erlauben, darf nicht als "Job" nur Lebenszeit verbrauchen.  
*Für die Automation* bedeutet diese ethische Regel, dass der Mensch für eine Vielfalt von Aufgaben heranzuziehen ist und ihm dabei jeweils die Gestaltungsmöglichkeiten nach 1. eröffnet werden sollen. Der Mensch muss aus seiner Arbeit Bestätigung erfahren.
3. Auch abhängige Arbeit ist so zu gestalten, dass Mitverantwortung für den Betrieb an Stelle einfachen Lohnempfangs gilt.  
*Für die Automation* bedeutet dies das Einbeziehen der Operateure in die Gestaltung der Prozesse und der Mensch-Prozess-Kommunikation sowie in die Definition der Prozessziele.
4. Arbeit ist gerecht zu verteilen unter Berücksichtigung von individuellen Fähigkeiten und Entwicklungsmöglichkeiten.  
*Für die Automation* ergibt sich daraus die Forderung nach einer fortwährenden Schulung und Weiterbildung der am Prozess beteiligten Menschen im Sinne der Erhaltung und Erweiterung ihrer individuellen Kompetenz.

Mehr Automation bedeutet zunehmende Verantwortung für den Menschen als der für den bestimmungsgemäßen Betrieb automatisierter Anlagen zuständige Instanz. Daraus folgern weitere technisch motivierte Richtlinien, die ebenfalls Teil einer Ethik der Automation sein sollten. Dazu gehören Forderungen und Anforderungen, wie sie vielfach diskutiert und in der Systementwicklung zunehmend aber keineswegs durchgängig berücksichtigt werden:

- *Systeme müssen robust ausgelegt werden gegenüber Fehlern des Menschen.*  
Da die Kette von Daten, Information, Wissen, Zielen und Aktionen auch für ein Automatisierungssystem gilt, beschränkt auf die vom Menschen definierten und implementierten Komponenten und Ziele, sind die Fehlerdetektions- und Reaktionsmöglichkeiten der Automatisierung entsprechend eingeschränkt. Dennoch ist diese Anforderung für die Realisierung der Anforderungen von de Quervain von großer Bedeutung.
- *Die Schnittstelle zwischen der Informationsverarbeitung und dem Menschen (siehe Bild 2) muss sicherstellen, dass der Mensch den ihm zugänglichen Informationen über den Prozess vertrauen kann.*  
Diese kritische Anforderung ist begründet durch den Abstraktionsgrad der Schnittstelle und damit durch die rezeptive und kognitive Distanz zwischen dem menschlichen Prozessoperator und dem technischen Prozess. Je mehr Verantwortung der Mensch übernehmen muss, um so wichtiger sind umfassende Informationen über den Prozesszustand, selbst wenn diese mit der heute verfügbaren Sensorik nicht zu beschaffen sind und ggf. durch Modelle oder sogar durch organisatorische Maßnahmen ersetzt werden müssen (z.B. Begehung des Betriebes durch zusätzliches Personal: Automatisierung schafft Arbeit).
- *Automatisierung soll den Menschen entlasten aber keinesfalls entmündigen.*  
Die Verantwortung für den bestimmungsgemäßen und sicheren Betrieb von (technischen) Prozessen liegt beim Menschen. Dieser Verpflichtung kann der Mensch nur dann gerecht werden, wenn er im Detail den aktuellen Zustand und Trend des Prozessablaufes nachvollziehen kann.

- *Bei den Nutzern der Automatisierung und Automation muss das Verständnis für die Grenzen dieser Technik und des Menschen geweckt werden.*  
Die Nutzer von (Dienst)Leistungen aus Prozessen, zu denen je nach Prozess auch Menschen und Einrichtungen außerhalb eines prozessbetreibenden Unternehmens gehören (z.B. Energieversorgung und -verteilung) müssen verstehen und ggfs. entsprechend ausgebildet werden, dass es in der Technik weder unbegrenzte Verfügbarkeit noch absolute Sicherheit geben kann. Damit ist die kritische Abwägung von Vorsorgeaufwendungen und Verfügbarkeitsrisiko angesprochen. Dazu gehört auch die Einsicht, dass Sicherheit vor Verfügbarkeit gehen muss.
- *Automatisierung und Automation sind immer wieder auf ihre Auswirkungen insgesamt zu überprüfen und kritisch zu bewerten.*  
Wenn Automatisierung den Betrieb eines Prozesses erst ermöglicht, ist die Frage zu beantworten, was bei Ausfall der Automatisierung geschieht und ob die Folgen eines solchen Ausfalles tolerabel sind hinsichtlich übergeordneter Ziele, die nicht allein die Automatisierung oder Automation betreffen (nicht beherrschte Prozesszustände / Unfälle jeder denkbaren Art).

Automatisierung hilft dem Menschen, die widerspenstige physikalisch-technische Welt für seine Ziele in Dienst zu nehmen. Sie ist damit ein unverzichtbarer Teil des menschlichen Lebens in seinen vielfältigen Ausprägungen. Die Zusammenarbeit der Automatisierungstechnik und des Menschen mit seiner Kreativität und Problemlösungsfähigkeit erfordert aus Sicht des Menschen Leitlinien für ihre Gestaltung und Durchführung. Diese ersten Gedanken zu einer Ethik der Automation sind sicher unvollständig und sollen eine entsprechende Diskussion anstoßen. Es ist aus Sicht des Autors aber zweifellos an der Zeit, die ungebremste Zunahmen der Komplexität technischer Systeme durch eine Diskussion über einen definierbaren Kern von Zielen im Sinne der Führung eines "guten Lebens" für die Gesellschaft zu beeinflussen.

## Quellen, Literatur

- [1] Weltingenieurtag 2000, Hannover: Diskussion zwischen dem Autor und Prof. M. Cooley (London) als Basis für diese Betrachtung
- [2] G. Geiser: Mensch-Maschine-Kommunikation. Oldenbourg Verlag, München, 1990
- [3] J. Rasmussen: Skills, Rules and Knowledge; Signals, Signs and Symbols, and Other Distinctions in Human Performance Models. IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-13, No.3, May/June 1983 pp. 257-266
- [4] R. Grimm: Bildschirme in der Prozesswarte. Richtlinien, Systemstrukturen und Tätigkeitsmerkmale, Vergleich zum Bürobereich; Verlag TÜV Rheinland, Köln 1983
- [5] Alfred de Quervain: Ruhe und Arbeit, Lohn und Eigentum. Ethik II,3, 1956



# **Automatisierungstechnik I**





# Proxykonzept zur Integration von intelligenten Feldgeräten in Steuerungssysteme

Christian Diedrich, Thomas Bangemann, Frank Naumann, Andreas Franke

Institut für Automation und Kommunikation Magdeburg – ifak, Steinfeldstr. 3, 39179 Barleben

## Abstract

Control system engineering and the instrumentation of industrial automation belongs to a very innovative industrial area as it is in the scope for cost reduction. That is why the costs for equipment and devices are already at a relatively low level. This trend is accompanied by a paradigm shift to digital processing in devices and digital communication between them. The proxy technology well known from the IT world is also used for field device integration via field bus into PLC control. This paper shows the principle of these proxies and describes the technology additionally using the ident system profile example for PROFIBUS.

**Keywords:** Decentral automation, instrumentation, UML, field device integration, field device profile, PROFIBUS

## 1 Einleitung

Wachsende Funktionalität von Feldgeräten und Antrieben und deren Kopplung über Kommunikationssysteme erfordern, dass die daraus resultierende Dezentralisierung auch eine funktionale Veränderung der anderen Systemkomponenten, wie z.B. Steuerungen und Leitsysteme, herbeiführt. Eine Veränderung der Steuerungs- und Leitsystemfunktionalität schlägt sich auch auf das Engineering und damit in den Softwarewerkzeugen nieder. Dieser Entwicklungstrend spiegelt sich in einer Menge von industriellen Initiativen und Aktivitäten auf dem Gebiet der nationalen und internationalen Standardisierung wider. Die folgenden Ausführungen konzentrieren sich auf das Zusammenwirken zwischen Steuerung und Feldgerät.

## 2 Das Stellvertreterkonzept (Proxy-Konzept)

Die angestammte Aufteilung der Steuerungsaufgaben zwischen Steuerungen und Feldgeräten hat sich durch den erweiterten Funktionsumfang von Feldgeräten verändert, die Funktionalitäten, wie z.B. Signalvorverarbeitung (vom Messumformer) und zum Teil sogar technologische Funktionen (von Antrieben) übernommen haben.

Die Verlagerung dieser Aufgaben hat zur Folge, dass das Steuerungsprogramm, das zuvor in einer Ressource (CPU einer Steuerung) abliefe und auch mit einem Programmierwerkzeug entwickelt worden ist, „zerrissen“ wird und in verschiedenen Geräten zur Abarbeitung kommt. Die Abarbeitungsreihenfolge wird nicht mehr innerhalb einer Steuerung organisiert, vielmehr wirken asynchron arbeitende Komponenten zusammen. Einige standardisierte Lösungen stehen schon seit längerer Zeit zur Diskussion ([IEC 61499, 2001] und [IDA, 2001]), sind aber noch nicht marktwirksam.

Um mit heutigen Programmiermitteln die Vorteile der Funktionsdezentralisierung/-verteilung nutzen zu können, gibt es in der Steuerung so genannte Stellvertreterbausteine (im Folgenden als Proxy bezeichnet), die, wie der Name sagt, dem Programmierer die

Feldgerätfunktionen stellvertretend bereitstellt. Ein Proxy ist ein Funktionsbaustein im Sinne der Programmiersprache Funktion Block Diagramm oder Structured Text [IEC 61131-3]. Die Eingänge des Proxy-Funktionsbausteins sind Variablen, die zum Feldgerät gesendet werden sollen und die Proxy-Ausgänge sind Variablen, die vom Feldgerät kommen. Es muss jedoch keine 1:1 Zuordnung zwischen Proxy-Funktionsbausteinvariablen und Feldgeräteparametern geben, da im Proxy selbst auch noch Verarbeitungsleistungen erbracht werden können (z.B. Datentypkonvertierungen, Maßeinheiten- und Skalierungsanpassungen).

Der benötigte Datenaustausch zwischen dem Proxy-Funktionsblock und den Gerätefunktionen geschieht über Kommunikationsdienste (Abb. 1). Diese Kommunikationsdienste werden durch so genannte Kommunikationsfunktionsbausteine realisiert, die Bestandteil der Proxies sind. Ein Proxy-Funktionsbaustein ist ein aus Kommunikationsfunktionsbausteinen und Proxytyp-spezifischen Algorithmen zusammengesetzter Baustein. Die Algorithmen wandeln die Feldgeräteparameter in die Funktionsbaustein-Ein- und Ausgänge um und organisieren die zeitliche und logische Abfolge der Interaktionen mit dem Feldgerät über den Feldbus. Der Proxy-Funktionsbaustein ist in die Ablauforganisation der Steuerung eingebunden, was dem Programmierer von der Kenntnis der Details der Funktionsabwicklung in den Feldgeräten und den benötigten Kommunikationsabläufen befreit.

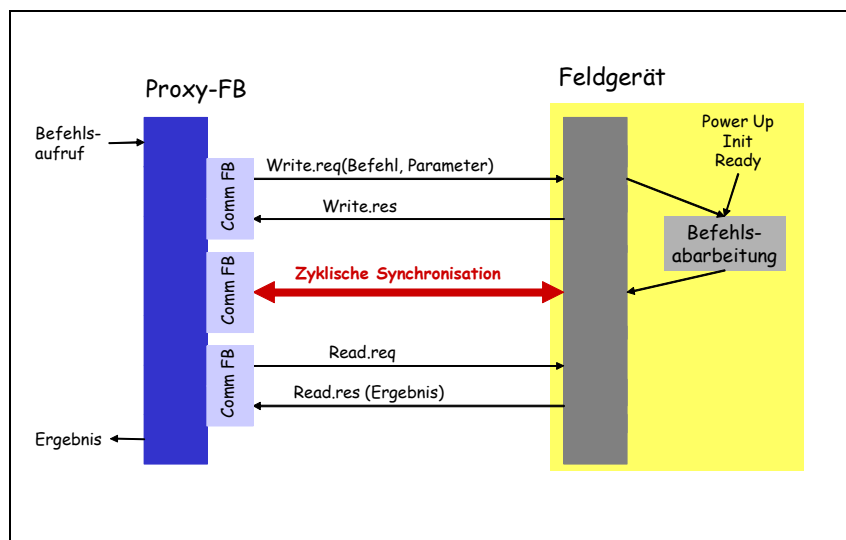


Abb. 1: Kommunikation zwischen Proxy-FB und Feldgerät

Der benötigte Datenaustausch geschieht über die Feldbuskommunikation zwischen Steuerung und Feldgerät sowohl über zyklische als auch azyklische Dienste, damit die ausgelagerten Funktionen parametrisiert werden können. Neu ist also, dass Steuerungen zusätzlich zum E/A-Abbild (meist realisiert durch den zyklischen Datenaustausch) Daten in der Steuerprogrammabarbeitung verwenden, die bei Bedarf, d.h. azyklisch gesendet oder angefordert werden. Für PROFIBUS ist unter Beteiligung von ifak-Mitarbeitern eine Richtlinie für die Entwicklung von Proxies im Allgemeinen und von Kommunikationsfunktionsbausteine [CFB, 2005] erarbeitet worden. Diese Kommunikationsfunktionsbausteine sind vereinheitlichte Treiber für das zyklische und azyklische Lesen und Schreiben von Feldgerätedaten, die zur Firmware oder zu den Standardbibliotheken der Steuerungen gehören. Sie sind Bestandteil der Proxies. Bei der Portierung der Proxies auf verschiedene Steuerungen, die die PROFIBUS-Kommunikationsbausteine unterstützen, ist somit keine

Anpassung an die Kommunikationseigenschaften notwendig. Die Proxies selbst sind in dem Maße einfach portierbar, wie die verwendeten Structured Text - Programmiersprachelemente den von PLCopen definierten Compliance-Stufen nicht überschreiten. Voraussetzung ist, dass die verwendeten Steuerungen entsprechende Compliance-Stufen unterstützen.

Proxies sind eine vom Feldgerätehersteller zusätzlich zu erbringenden Leistung, die eine Folge der Dezentralisierung ohne Standard für ein echtes verteiltes System ist. Feldbusprofile können die Kernfunktionalität dieser Proxies bilden, was den Aufwand bei der Erstellung minimiert. Proxies bilden aber auch eine gute Basis, kundenspezifische Funktionalitäten zunächst in der Steuerung zu realisieren, bevor diese in einer neuen Geräteversion fest in die Feldgeräte implementiert werden. Sie bieten Feldgeräteherstellern weiterhin die Option, ihren Kunden eigene, herstellereigene Features auf einer übergeordneten Plattform (Master SPS) anbieten zu können (Komfort als Marketingargument).

### **3 Anwendung des Proxy-Konzepts bei Identifikationssystemen**

#### **3.1 Das Ident-Profil für PROFIBUS**

Identifikationssysteme auf Transponderbasis und Barcodeleser sind aus Anwendungen wie z.B. der Produktverfolgung während des Produktionsprozesses, der Erkennung eines individuellen Objektes oder der Speicherung von Daten direkt am Produkt, heute nicht mehr wegzudenken. Am Markt findet man eine Vielzahl von Lösungen unterschiedlichster Hersteller. Leider sind diese Lösungen untereinander nicht kompatibel. Logische Konsequenz ist ein Mehraufwand bei Bedienung und Integration in Steuerungssysteme. Weder die Luftschnittstelle noch die Übertragung der Daten zur SPS finden in standardisierter Form statt. Es liegen herstellereigene, geschlossene und somit nicht interoperable Lösungen, meist auf RS232-Basis, vor. Betrachtet man jedoch den Trend in der Automatisierungstechnik, werden immer mehr Feldgeräte über Feldbusschnittstellen miteinander vernetzt. Der Einsatz eines einheitlichen Kommunikationsprotokolls allein reicht jedoch nicht aus, um die Interoperabilität zwischen den Systemkomponenten zu gewährleisten. Optionale Eigenschaften sind einheitlich einzusetzen und Absprachen bzgl. Syntax und Semantik von Daten zwischen den Geräteherstellern, den Steuerungs- sowie den Leit- und Engineeringssystemen zu treffen. Auf applikativer Ebene werden diese Festlegungen in sogenannten Profilen getroffen, wie im Folgenden am Beispiel von PROFIBUS dargestellt werden soll.

#### **3.2 PROFIBUS als effektives Transportmittel**

Aktuelle Basis für die effektive PROFIBUS-Kommunikation ist das PROFIBUS-DP Protokoll. PROFIBUS-DP wird meist für die Kommunikation zwischen einem zentralen Gerät (Mono-Master-System) und mehreren verteilten Ein-/Ausgabeeinheiten verwendet. Das bedeutet, dass ein Master-Gerät über PROFIBUS-DP mit den verteilten Peripheriegeräten (Slaves) vernetzt ist. Die in klassischer Architektur aufwendige Parallelverkabelung zwischen der SPS und den Geräten der dezentralen Peripherie wird durch eine serielle Busleitung ersetzt, über die ein zyklischer Prozessdatenaustausch erfolgt.

PROFIBUS DP ist der Überbegriff für die 3 Protokollvarianten DP-V0, DP-V1 und DP-V2. Zunächst wurde DP-V0 für den rein zyklischen Datenverkehr entwickelt. Die Weiterentwicklungen in der Automatisierungstechnik (die Geräte werden komplexer und müssen häufig konfiguriert und parametrisiert werden) weckten Bedarf nach mehr

Bedienkomfort. Das rein zyklische DP-Protokoll (DP-V0) ist jedoch nur bedingt für Parametrierzwecke geeignet, da die Kodierung von Parameterinformationen in Nutzdatentelegrammen zu einem hohen Telegramm-Overhead führen würde. Ein Ausweg aus dieser Situation wurde durch die Entwicklung der Version DP-V1 realisiert. Mit DP-V1 ist es möglich, parallel zum zyklischen Prozessdatenaustausch azyklisch Telegramme vom gleichen bzw. einem anderen Master zu senden. DP-V2 ist eine Weiterentwicklung, die z.B. Slave-Querverkehr, Isochron-Mode oder Download unterstützt.

Darüber hinaus werden in sogenannten Profilen Vereinbarungen zwischen Herstellern für Funktionalitäten oberhalb der Schicht 7 getroffen. Profile sind Festlegungen, die Feldgerätfunktionen (z.B. messen, stellen), verbunden mit ihren Variablen und Parametern so beschreiben, dass auf diese herstellerunabhängig über den Feldbus zugegriffen werden kann. Sie definieren weiterhin das zu nutzende Subset des Kommunikationsprotokolls. Ein solches Anwendungsprofil ist das Profil „Identification Systems“ [PNO, 2005].

Das Ident-Profile, das unter technischer Leitung von ifak-Mitarbeitern entstanden ist, ist das erste Profil einer neuen Profildeneration, die nicht die Details der Feldgeräte im Einzelnen beschreibt, sondern den oben beschriebenen Typ eines Stellvertreterbausteins im Steuerungsprogramm mit einem standardisierten Interface (d.h. definierten Funktionsblockein- und -ausgängen) und den unterlagerten Kommunikationsmechanismen zwischen dem Master und dem Slave.

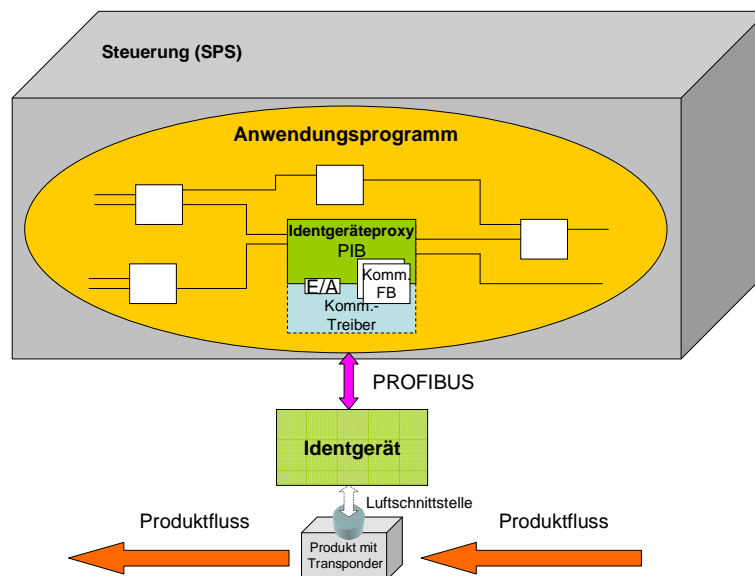


Abb. 2: Zusammenwirken zwischen Identgerät und seinem Proxy im Steuerungsprogramm

Bei der Programmierung des Steuerungsprogramms kann so aus der Bibliothek ein Identgerätebaustein herausgenommen und in das Anwendungsprogramm eingebunden werden, obwohl die wesentlichen Identifizierungsfunktionen im Identgerät erbracht werden. Für die Kommunikation zwischen dem Master und dem Slave kommen der zyklische Verkehr nach DPV0 (in Abb. 2 sichtbar durch die E/A Schnittstelle zwischen Anwendungsprogramm und Kommunikationstreiber) und azyklische Lese- und Schreibdienste lt. DPV1 (in Abb. 2 sichtbar durch die Kommunikations-Funktionsblock (Komm.-FB) Schnittstelle zwischen Anwendungsprogramm und Kommunikationstreiber) zum Einsatz. Abb. 3 zeigt diese Struktur im Detail.

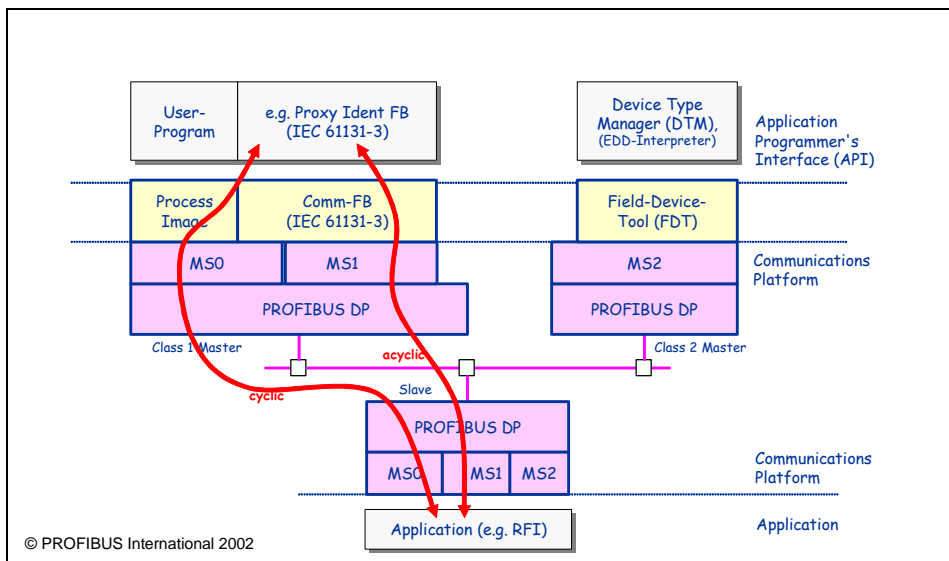


Abb. 3: Kommunikationsarchitektur

### 3.3 Einfache Handhabung durch Profilstellungen

Voraussetzung für die Interoperabilität der Systeme, d.h. hier der Steuerungen und Identsysteme, sind Absprachen über das Kommunikationsgeschehen und die zu verwendende Programmiersprache zu treffen. Es wurde deshalb im Profilarbeitskreis beschlossen, einen Baustein (Arbeitstitel: **Proxy Ident Baustein – PIB**) zu definieren, der:

- entsprechend IEC 61131-3 spezifiziert wird,
- eine definierte Schnittstelle zu den im Rahmen der PNO-Arbeiten definierten Kommunikations-Funktionsbausteinen (Comm-FB's) sowie
- eine definierte Schnittstelle zum Steuerungsprogramm hat.

Die Comm-FB's, standardisiert in einer PROFIBUS Guideline (siehe Spezifikation auf den PROFIBUS-Seiten im Internet) und die Verbreitung von IEC 61131-3 Sprachen (vor allem "Structured Text") sind die Voraussetzungen für die Interoperabilität des Bausteins. Der Anwendungsprogrammierer kann sich somit auf die automatisierungstechnische Aufgabenstellung konzentrieren und muss kein PROFIBUS-Spezialist sein.

Hauptaufgabe der Kommunikation zwischen Identgerät und Steuerungsprogramm ist der Austausch von Datenpaketen, die einige Bytes (z.B. Barcodes) oder -zig kByte (z.B. Transponder mit zum Teil 32 kByte) betragen können. Während bei anderen Profilen die Definition der Dateninhalte im Vordergrund steht, werden die Übertragung der Datenpakete (Dateien oder Teile von Dateien) sowie die Dateistruktur zunächst außer Acht gelassen. Die im Profil spezifizierten Mittel beziehen sich auf Transportmechanismen für „Datencontainer“. Der PIB zerlegt bzw. setzt größere Datenpakete in bzw. aus Teilpaketen intern zusammen. Das Steuerungsprogramm übergibt oder empfängt nur ganze Datenpakete, die auch nur von diesem verarbeitet werden. Für den PIB sind die Datenpakete transparent. Der PIB hat folgende Funktionalitäten:

- Zerlegen und Zusammensetzen von Datenblöcken,
- Flusskontrolle für die sequenzielle Übertragung der zerlegten Datenblöcke / Synchronisation nach einem Fehlerfall,
- Behandlung von Übertragungsfehlern auf Anwendungsebene,
- Identifikation des Ident-Kanals.

Die beschriebenen Anforderungen von Transpondern an Automatisierungssysteme sind aus Sicht des Kommunikationsgeschehens eher untypisch. Die markierten Objekte werden von den Schreib-/Lesestationen ("Interrogator") in relativ großen Zeitabständen (Sekunden bis Minuten) bearbeitet. Zur Gewährleistung einer großen Performance des Gesamtsystems erfordert diese Eigenschaft besondere Sorgfalt bei der Koordination der zyklischen und azyklischen Kommunikation. Demnach wird im zyklischen Betrieb lediglich ein kurzes Status- bzw. Quittungstelegramm (16 Bit) ausgetauscht, welches im Wesentlichen über den Bearbeitungszustand eines Befehls im Interrogator berichtet. Die Befehle an sich sowie Befehlsantworten und Daten werden azyklisch in die Steuerung übertragen (Das Schreiben von Daten erfolgt entsprechend.). Durch die zyklische Statusübertragung werden azyklische Dienste nur dann aufgerufen, wenn auch tatsächlich Daten vorliegen. Bei der Parametrierung der Buszyklen ist zu gewährleisten, dass in jedem Buszyklus eine „Reservezeit“ für azyklischen Datenverkehr verbleibt.

Durch diese Vorgehensweise ergeben sich eine Reihe von Vorteilen

- es erfolgt eine geringe Belastung des zyklischen Datenverkehrs (Schnelle E/A-Baugruppen behalten ihre Reaktionszeiten, kurze PROFIBUS-Zykluszeiten auch bei vielen PROFIBUS-Slaves),
- optimaler Datendurchsatz,
- unnötiger zyklischer Datenverkehr wird vermieden, d.h. dass nur die benötigten Daten übertragen werden.

### 3.4 Profilentwicklung und -implementierung

Auch bei der Entwicklung des Profils wurden neue Wege beschritten. Neben den Interfacebeschreibungen für den Identgerätebaustein besteht dieser aus mehreren interagierenden Zustandsmaschinen. Im Detail sind die Zustandsmaschinen für die Bausteinverwaltung, die Status- und Steuerwortsteuerung, die Befehlsabarbeitung mit der Paketierung der Daten und der azyklischen Kommunikations-FBs zu koordinieren. Diese komplexe Aufgabe ist ohne Werkzeug kaum in benötigter Qualität zu entwerfen. Deshalb wurde spezifikationsbegleitend ein Modell, basierend auf der Unified Modeling Language (UML), mit dem Softwarewerkzeug Rhapsody in C erstellt und simuliert. Die Simulationsergebnisse gaben hilfreiche Inputs für die Diskussionen auf den Arbeitskreissitzungen.

Die Spezifikation beinhaltet sowohl den Identgerätebaustein (PIB) als auch die benötigte Identgerätefunktionalität als Gegenstelle, die als PIB\* bezeichnet wird. Beide Bausteine bestehen aus den oben beschriebenen komplexen Zustandsmaschinen, die als UML– State Chart vorliegen. Zur simulativen Erprobung wurden in C++ Oberflächen geschrieben, die die Bedienung des PIB und des PIB\* ermöglichen (Nachbildung der externen Bausteinparameter). Die PROFIBUS-Kommunikation zwischen Master und Slave erfolgt im weiteren Entwicklungsprozess, indem jeweils die PROFIBUS-Master und Slave-Kartentreiber

in UML eingebunden werden, über die auf die Master- und Slavekontroller zugegriffen wird. Beide können sich in einem oder in verschiedenen PC's befinden.

Im Ergebnis liegen sowohl der PIB als auch der PIB\* als erprobte UML-Spezifikation und auch als C Quellcode vor. Der PIB wird manuell in IEC 61131-3 Structured Text umgesetzt und kann auf Steuerungen, die die benötigten Kommunikationsbausteine besitzen, wie z.B. einer Simatic S7-300 ablaufen. Der aus dem Rhapsody in C generierte C Code des PIB\*, der das Profil-Interface für den Slave darstellt, kann als Treiber in die PROFIBUS-Slave-Anschaltung von Identifikationssystemen eingebunden werden. Dadurch ist die Profilintegration in die Feldgeräte wesentlich vereinfacht worden. Die Hersteller können auf eine validierte Software zurückgreifen.

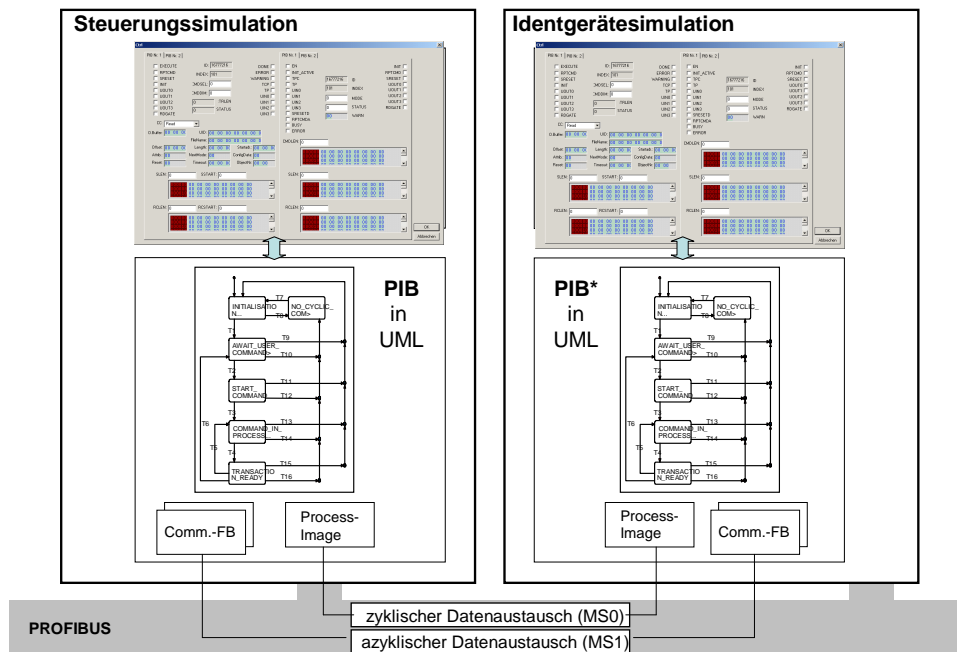


Abb. 4: UML-basierte Simulation des Identprofils

## Ausblick

Das hier beschriebene Proxy-Konzept wird bei PROFIBUS neben den Identprofil auch für Waagen und Dosiersystem eingesetzt. Es ist zu erwarten, dass dieses Konzept auch für weitere Geräteintegrationen umgesetzt wird.

## Literatur

- [CFB, 2005] PNO Specification PROFIBUS and PROFINET IO Draft Guideline: IEC 61131 Communication Function Blocks on PROFIBUS DP and PROFINET IO. V1.9. PNO Karlsruhe August 2005. Order No.:2.182.
- [IDA, 2001] IDA: IDA – the Internet of Automation Technology. White paper V1.0 April 2001. [www.ida-group.org](http://www.ida-group.org).
- [IEC 61131-3, 1992] IEC 1131-3: Programming Language for Programmable Controllers, Geneva 1992.
- [IEC61499, 2004] IEC 61499: Function Blocks for industrial-process measurement and control systems – Part1 and 2. Geneva 2004.
- [Neumann, 2001] Neumann, P., Simon R., Diedrich, Ch., Riedl, M. (2001). Field Device Integration. 8th IEEE International Conference on Emerging Technologies and Factory Automation. ETFA 2001, Proceedings pp.63-68, Antibes.
- [PNO, 2005] PROFIBUS Profile Guideline „Identification Systems – Proxy Ident Function Block“ Draft Version 1.0, März 2005



# Vorschlag für ein Prozessinformationsdisplay zur Übersichtsdarstellung komplexer Prozesszustände

Dipl.-Ing. (FH) Knut Meißner, Phone: ++49.3943-659-320, Fax: ++49.3943-659-399, E-Mail: kmeissner@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

Prof. Dr.-Ing. Hartmut Hensel, Phone: ++49.3943-659-313, Fax: ++49.3943-659-399, E-Mail: hhensel@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

## Abstract

Aus den steigenden Anforderungen an die Bediener moderner technischer Anlagen entsteht der Bedarf an neuen, auf den Menschen und seine Fähigkeiten und Eigenschaften besser abgestimmten Benutzungsschnittstellen. In diesem Beitrag werden die bei der Untersuchung klassischer Visualisierungsverfahren ermittelten Anforderungen in einem neuen Visualisierungsverfahren umgesetzt. Um dabei ein hohes Maß an Flexibilität zu erreichen, wurden verschiedene Werkzeuge entwickelt. Diese gestatten zum einen eine schnelle Anbindung an Standardprogramme für den Entwurf von 2-D- und 3-D-Welten als auch eine einfache und flexible Anbindung an verschiedene technische Prozesse über Standardschnittstellen. Das hier vorgestellte Visualisierungskonzept beruht auf einer kognitiv-szenischen Darstellung der Prozesswerte in einer 3-D-Welt, die die mentalen Fähigkeiten des Bedieners unterstützen und so Informationsüberflutungen verhindern soll.

**Keywords:** Mensch-Maschine-Interaktion, Visualisierung, 3-D-Methoden, Prozessleittechnik

## 1 Einleitung

Die aktuelle technische Entwicklung in der Prozessindustrie bewirkt eine deutliche Erhöhung der Anlagenkomplexität und somit eine Steigerung der Anforderungen an die Bediener. Der gestiegene Automatisierungsgrad, die verbesserte Informationserfassung in den Anlagen und die gleichzeitige Verringerung des Bedienpersonals führen zu einem stetigen Anwachsen der durch den einzelnen Bediener zu überwachenden Informationsmenge [8]. Berücksichtigt man die daraus resultierende und in [6] beschriebene Überlastung der kognitiven Fähigkeiten des Bedieners, ergibt sich insbesondere in Alarm- und Stresssituationen ein gefährliches Fehlerpotenzial in den Bedieneraktionen. In der Fachliteratur wurden bereits verschiedene Lösungswege diskutiert, die diesem Problem begegnen sollen. Sie reichen auf der einen Seite von einer verbesserten Visualisierung wie dem Massendatendisplay [5] über die virtuellen Prozesselemente [10] bis hin zur 3-D-Visualisierung [4 und 10], die den Benutzer entlasten soll. Auf der anderen Seite wurden komplexe Expertensysteme [3] entwickelt, welche den Bediener mittels Methoden der künstlichen Intelligenz, mathematischer Prognosemodelle und gespeicherten Prozesswissens aktiv bei der Prozessführung unterstützen sollen. Die Menge der in unterschiedlichen Anlagentypen zu überwachenden Prozessgrößen und die Anzahl der eingesetzten Prozessbilder zeigt eine beispielhafte Studie von Alty et al. [1] in Tabelle 1. Zum anderen wuchs die Zahl der zu überwachenden Prozessbilder pro Anlage und die Menge der Prozessvariablen pro Prozessbild in den letzten 15 Jahren stetig an. Nach einer Studie von Völkel [9] in einer typischen Leitwarte ist ein Anlagenfahrer in einer modernen Anlage für die Kontrolle von bis zu 1500 Prozessleitstellen verantwortlich.

Tabelle 1: Verhältnis zwischen Prozessgrößen und Prozessbildern

Anlagentyp	Anzahl Prozessgrößen	Anzahl Bilder	Prozessgrößen pro Bild
Energieversorgungsnetze	100000 – 200000	100 – 200	100 – 200
Prozessindustrie	10000 – 20000	ca. 100 R&I Bilder und ca. 300 Standardbilder	< 200
Kraftwerke	2000 – 10000	ca. 100 R&I Bilder und ca. 300 Standardbilder	< 300
Walzwerke	< 100	1 – 2	5 – 10
Flugzeug	100 – 200	5 – 100	5 – 30

Durch eine angepasste Form der Prozessdatenvisualisierung, die auf Erkenntnissen der Kognitionswissenschaft beruht, werden die kognitiven Fähigkeiten der Benutzer besser unterstützt und genutzt. Hierbei handelt es sich um einen kognitiv-szenischen Ansatz, der dem Bediener das Erkennen und Klassifizieren von kritischen Prozesszuständen oder Prozessenden erleichtert. Im zweiten Teil werden die Ergebnisse einer Anforderungsanalyse an Visualisierungssysteme in der Prozessleittechnik vorgestellt. Im dritten Abschnitt wird dann ein Lösungsansatz zur kognitiv-szenischen Darstellung von komplexen Prozesszuständen diskutiert. Der vierte Abschnitt beschreibt die beim praktischen Einsatz mit diesem neuartigen System gewonnenen Erfahrungen und in Abschnitt fünf werden kurz einige Ergebnisse der Evaluation des Visualisierungsverfahrens präsentiert. Den Abschluss bilden eine Zusammenfassung der Ergebnisse und ein Ausblick auf die weiteren Arbeiten.

## 2 Analyse der traditionellen Visualisierungsmethoden

Bei den klassischen Visualisierungsverfahren in der Prozessleittechnik wie Grafik- und Gruppenbild muss der Anlagenfahrer eine Vielzahl alphanumerisch und grafisch dargestellter Messwerte beobachten und bewerten. Dieser Beobachtungsvorgang ist zum einen durch die große Monotonie bei der standardmäßigen Überwachung im Normalbetrieb und zum anderen durch eine sehr hohe und einseitige kognitive Belastung des Bedieners in Stress- oder Ausnahmesituationen [7] gekennzeichnet. Solche Ausnahmesituationen entstehen durch Fehlfunktionen der Anlage oder bei Sonderbetriebsarten wie Wartungsvorgängen und Wiederinbetriebnahmen. Stresssituationen führen nachweislich zu einer erhöhten Fehlerrate bei den Bedienoperationen, die wiederum das Stresspotenzial einer Situation weiter erhöhen und somit die Wahrnehmung des Bedieners negativ beeinflussen. Die kognitiven Fähigkeiten des Unterbewusstseins zur zwei- und dreidimensionalen Mustererkennung werden in den klassischen Prozessgrafiken nicht oder nur kaum genutzt. Das Massendatendisplay ist ein an der TU Clausthal entwickeltes Visualisierungskonzept, das die Signale und Prozessgrößen einer Anlage oder eines Anlagenteils in Form von Strichen abbildet. Das daraus entstehende Muster ist vom Bediener leicht erkennbar und auswertbar. Abweichungen von den Sollwerten werden durch eine Winkeländerung der Striche bzw. eine Vergrößerung der Strichstärke visualisiert. Zu dieser Darstellung ist weiter anzumerken, dass durch die räumliche Zuordnung von Prozessgröße zum physikalischen Ort in der Anlage

(Prozesstopologie) bereits zwei Darstellungsdimensionen für die Visualisierung eingesetzt werden. Eine abstraktere Möglichkeit der Visualisierung des Prozesszustandes mittels geeigneter Ikone wird nach einer Anforderungsanalyse vorgestellt, die im nächsten Abschnitt erfolgt.

### **3 Anforderungsanalyse**

Aus den in Tabelle 1 genannten Zahlen ergeben sich für die Entwicklung einer an den Menschen angepassten Darstellungsform komplexer Prozesszustände Anforderungen, die im Weiteren aufgeführt und diskutiert werden. Dabei ist zwischen den Anforderungen der Bediener und des Einsatzgebietes, in diesem Fall das Gebiet der Prozessleittechnik, zu unterscheiden. Im zweiten Schritt muss dann versucht werden, die beiden Anforderungsprofile bestmöglich zu kombinieren. Die wichtigsten Anforderungen der Prozessleittechnik an ein Visualisierungssystem sind nach dieser Untersuchung eine präzise Darstellung der Informationen und Botschaften, eine geringe Interpretationsmöglichkeit, eine Darstellung aller Ereignisse, die ein Eingreifen erfordern, Echtzeitfähigkeit und die Darstellung von Überblick- und Detailinformationen.

Für den Bediener sind die folgenden Punkte als Hauptanforderungen ermittelt worden. Die Informationen sollten benutzer- und situationsangepasst dargestellt, ein einfaches Erkennen von Mustern und Strukturen innerhalb der Informationen sollte zur Verminderung des Suchaufwandes nach Informationen unterstützt, zur Bildung von Zusammenhängen sollten Farben eingesetzt und durch dynamische Vorgänge wie Blinken und Bewegung sollte die Aufmerksamkeitssteuerung unterstützt werden.

Anhand dieser beiden Anforderungsprofile wurde das im nächsten Abschnitt vorgestellte 3-D-Prozessinformationsdisplay entworfen.

### **4 Aufbau des 3-D-Visualisierungssystems**

Die hier vorgestellte kognitiv-szenische Visualisierung beruht auf einer symbolischen Darstellung der Prozesswerte durch Ikone. Hierdurch soll die bereits angesprochene kognitive Belastung reduziert werden. Das Erkennen der Situation soll dabei vom Bewusstsein in das Unterbewusstsein verschoben werden und der Bediener soll nicht nur einzelne Fehler, sondern Situationsklassen erkennen. Wie in Bild 3a abgebildet, besteht ein Ikon aus einem äußeren Ring, einer inneren Anzeigefläche, einem Prognosezeiger, einem optionalen Trend und einem Bezeichner. Der äußere Ring ermöglicht dem Bediener eine farbliche Zuordnung eines Ikon zu einer Prozessgröße. Die innere Anzeigefläche signalisiert dem Bediener die Dringlichkeit eines evtl. Eingriffes in den Prozess. Im Sollzustand ist die Anzeigefläche transparent. Dies ermöglicht auch die Sicht auf dahinter liegende Ikone. Mit steigender Abweichung des Istwertes einer Prozessgröße vom Sollwert färbt sich die Anzeigefläche stärker rot. Zusätzlich zur Farbinformation wurde noch ein szenischer Aspekt in die Visualisierung integriert.

Aufgrund der im praktischen Einsatz gesammelten Erfahrungen wird die innere Anzeigefläche nicht mehr wie in früheren Versionen durch einen Zylinder dargestellt, sondern, wie in Abbildung 3b dargestellt, durch einen Ellipsoid. Dadurch kann die Farbe der inneren Anzeigefläche auch im Falle eines ungünstigen Betrachtungswinkels noch eindeutig erkannt werden. Die Dringlichkeit des Eingriffes wird nicht nur durch das Einfärben der Anzeigefläche visualisiert, sondern zusätzlich durch eine Vergrößerung aufgrund des In-den-Vordergrund-Rückens des Ikon bewirkt.

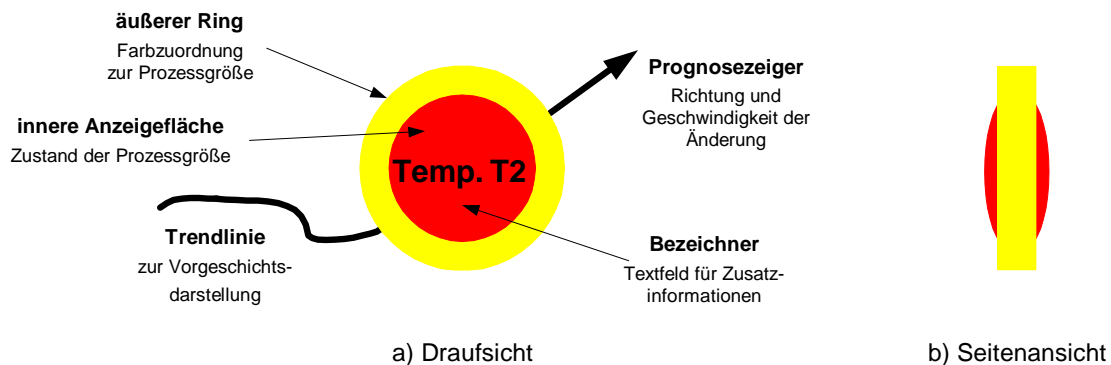


Abb. 1: Struktur eines Visualisierungssikons

Ein Wert, der einen dringenden Eingriff erforderlich macht, rückt in den Vordergrund der Visualisierung und entspricht somit der kognitiven Erwartung und Erfahrung des Bedieners aus der realen Welt. Dies wurde durch die 3-D-Darstellung des Prozessraumes erreicht. Die Visualisierung einer Prognose soll es dem Bediener ermöglichen, die Bewegungsrichtung und Bewegungsgeschwindigkeit der Prozessgrößen im Prozessraum besser einzuschätzen. Hierzu wurden lineare und nicht lineare Prognosemodelle der zugrunde liegenden Zeitreihenanalyse untersucht, um dem Bestreben des Bedieners nach einer Antizipation linearer Zeitverhaltens des weiteren Prozessverlaufes entgegenzuwirken. Berührt der Benutzer ein Icon mit dem Mauszeiger, wird zusätzlich ein Bezeichner mit dem Namen des Prozesswertes eingeblendet. Abbildung 4a zeigt mehrere dieser Ikone in einem beispielhaften 3-D-Prozessraum. Alle Prozessgrößen liegen innerhalb der Spezifikation, somit sind alle Ikone in der Nähe des Ursprungs, d. h. in der Bildmitte positioniert. Durch diese Art der Anordnung wird zusätzlich eine visuelle bzw. grafische Kompression erreicht, welche die Bildung von Informationspaketen (Chunks) im Arbeitsgedächtnis erleichtert. Die Achsenbeschriftungen und Darstellungsfunktionen der einzelnen Ikone werden über ein Zusatzprogramm konfiguriert und alle Eingangsgrößen werden auf einen Darstellungsbereich normiert. Im Gegensatz zu Abbildung 4a zeigt Abbildung 4b einen Fehlerzustand. Die Einfärbung der Anzeigefläche und das In-den-Vordergrund-Rücken der Ikone mit einer hohen Dringlichkeit (Temp. T2 und Level Vorlauf) sind gut erkennbar. Die Skalen und Gitter des 3-D-Raumes dienen dem Quantifizieren der analogen Anzeigen und sollen dem Bediener lediglich ein orientierendes Ablesen ermöglichen, so wie in der VDI/VDE 3699 Blatt 2 gefordert.

Aus diesem Grund wird eine normierte Darstellung, bei der die Skaleneinteilung genau diesem orientierenden Ablesen dient, eingesetzt. Diese Art der Darstellung bietet den Vorteil einer generischen Visualisierung unterschiedlicher Prozessgrößen und physikalischer Einheiten in einer Darstellung. Nicht der Wert der Prozessgröße an sich ist hier interessant, sondern nur die transformierten Größen. Die z-Achse wird ebenfalls nicht über eine physikalische Größe aufgespannt. Sie symbolisiert die Dringlichkeit, d. h. die Notwendigkeit eines Eingriffes durch den Bediener. In diesem Beispiel wurde auf der x-Achse die Abweichung des Prozesswertes vom Sollwert und auf der y-Achse die Abweichung vom Mittelwert abgetragen.

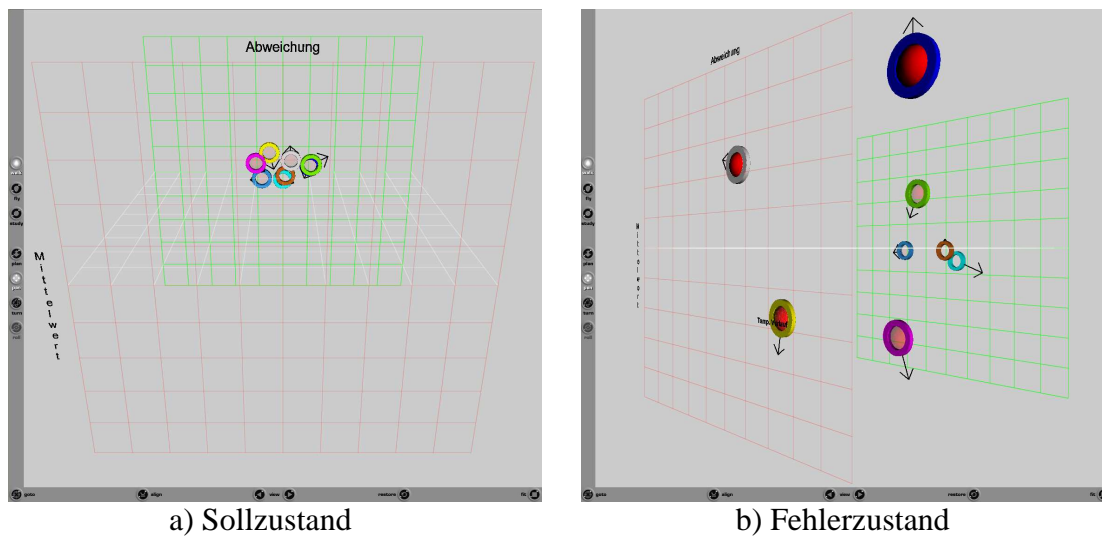


Abb. 2: 3-D-Prozessinformationsdisplay

Durch diese Weiterentwicklung des bereits in früheren Veröffentlichungen vorgestellten 3-D-Visualisierungssystems und insbesondere den erweiterten Aufbau des Ikon sowie die modifizierte Darstellung der Dringlichkeit im 3-D-Raum konnte eine Verbesserung beim Erkennen und Bewerten der Ikone erreicht werden. Zusätzlich wurden neben der linearen Dringlichkeit auch Untersuchungen mit logarithmischen und exponentiellen Dringlichkeitsfunktionen durchgeführt. Die Vorteile dieser beiden komplexeren Dringlichkeitsfunktionen liegen zum einen in der Möglichkeit, störendes Prozessrauschen durch das Unterdrücken kleinerer Prozesswertschwankungen (exponentielle Dringlichkeit) herauszufiltern. Zum anderen können mittels der logarithmischen Dringlichkeitsfunktion bereits kleine Prozesswertabweichungen stark verdeutlicht werden. Nachteilig sind die komplexeren funktionalen Zusammenhänge und die daraus resultierende höhere Belastung des Bedieners beim Erkennen der Dringlichkeit sowie eine steigende Überdeckung der Ikone im 3-D-Raum. Das Problem der Darstellung eines Ereignisschwall konnte in dieser Darstellungsform ebenfalls durch die Verwendung der grafischen Dringlichkeitsfunktion gelöst werden. Jeweils die Prozessgröße, die die höchste Dringlichkeit im Falle des Ereignisschwall besitzt, wird auch im Vordergrund der Darstellung präsentiert. Dadurch ist der Bediener in der Lage, die einzelnen Eingriffe nach ihrer Dringlichkeit vorzunehmen. Allerdings ist diese Form der Darstellung kein Ersatz für eine Alarmliste, da hier keine Zeitstempel für das Aufkommen der Alarme und die Bedieneingriffe protokolliert werden können.

## 5 Evaluation des Visualisierungssystems

Im Anschluss an die Analyse der bestehenden Visualisierungsverfahren und den Entwurf wurde die Funktionsweise im praktischen Einsatz getestet, um die theoretische Herleitung der neuartigen Mensch-Maschine-Schnittstelle zu überprüfen. Diese Überprüfung fand sowohl unter realen Betriebsbedingungen in einem Werk zur Kalkherstellung als auch unter Laborbedingungen statt. Für die statistische Auswertung der Befragungen wurde eigens ein Fragebogen entwickelt. Dieser Fragebogen basiert auf dem an der University of Maryland entwickelten *Questionnaire For User Interaction Satisfaction* (QUIS) in der Version 7, der von einem interdisziplinären Forscherteam erarbeitet wurde. Für die Anwendung innerhalb

dieses Projektes wurde der Fragebogen in einem kooperativen Projekt mit der University of Maryland ins Deutsche übersetzt. Da es sich bei *QUIS* um einen Fragebogen zur Überprüfung allgemeiner Benutzungsschnittstellen handelt, wurde er durch weitere Fragenkomplexe ergänzt, die speziell auf Benutzungsoberflächen im leittechnischen Umfeld zugeschnitten sind. Trotz der Anwendung eines vielfach überprüften Fragebogensystems, wie es *QUIS* darstellt und größter Sorgfalt bei der Befragung und Auswertung bergen die ermittelten Ergebnisse Unsicherheiten. Diese resultieren daher, dass die Ergebnisse von Probandenversuchen durch weitere, nicht nachvollziehbare Variationen von Einflussfaktoren beeinflusst werden können [2]. Dies sind zum einen die Vorbildung der für die Befragung herangezogenen Probanden, aber auch die Art der dargestellten Informationen und die Form der Untersuchung und der Darstellung selbst. Durch den Einsatz eines Prozesssimulators wurde versucht, auftretende Störungen oder Unterschiede zwischen den einzelnen Befragungen zu minimieren und evtl. unvorhergesehene Prozessabweichungen zu verhindern, damit alle Probanden unter gleichen Bedingungen befragt werden konnten.

Eine weitere Schwierigkeit besteht in der Auswahl geeigneter Probanden für die Befragung. Hierbei stellt sich insbesondere die Frage, inwieweit die Ergebnisse einer Befragung von Studenten auf das Verhalten von realen Operateuren schließen lassen. Aufgrund der bekannten und verbreiteten Skepsis und Befangenheit der Bediener solcher Anlagen wurden die Tests nicht nur mit ausgebildeten Operateuren, sondern parallel mit 20 Studenten der Fachrichtungen Automatisierungstechnik und Ingenieurinformatik durchgeführt. Durch diese Maßnahme sollte sowohl die Validität gesichert als auch die Objektivität der Befragung erhöht werden.

Neben der Evaluation im Labor wurde das 3-D-Visualisierungsverfahren auch innerhalb einer realen Anlage zur Herstellung von Branntkalk getestet. Die Ankopplung des 3-D-Visualisierungsverfahrens an den realen Prozess wurde über einen OPC-Server der Firma INAT realisiert. Dieser OPC-Server diente ebenfalls zur Kommunikation des bereits bestehenden SCADA-Systems mit der unterlagerten Steuerung. Im Rahmen dieser Evaluationen wurden bis zu 45 Prozesswerte innerhalb eines 3-D-Prozessinformationsdisplays dargestellt. Die Anzahl der darstellbaren Prozessgrößen wurde dabei nicht durch das Visualisierungsverfahren, sondern durch die Leistungsfähigkeit der eingesetzten VRML-Komponente begrenzt. Für den Industrieinsatz ist in jedem Fall die Umsetzung mittels einer leistungsstärkeren 3-D-Grafikkomponente notwendig.

## **6 Ergebnisse der Evaluation**

Aufgrund des Umfangs der Untersuchungen und Auswertungen – insgesamt 92 Fragen zu zehn Fragenkomplexen – können an dieser Stelle nur die wichtigsten Ergebnisse präsentiert werden. Die gesamten Ergebnisse sind Gegenstand einer weiteren Veröffentlichung, die in nächster Zeit erscheinen wird. Bei der durchgeführten Befragung wurden die Möglichkeiten unterschiedlicher Darstellungsformen für Prozessdaten miteinander verglichen. Dabei wurde das neu entwickelte 3-D-Prozessinformationsdisplay den traditionellen Darstellungsformen Grafikbild, Gruppenbild und Kurvendarstellung gegenübergestellt. Im Rahmen der Evaluation sollten im Wesentlichen die Umsetzung der in Abschnitt 2 genannten Anforderungen und die folgenden Fragestellungen untersucht werden:

- Wie wird die Nützlichkeit der Merkmale des Ikon und des 3-D-Raums bewertet?
- Ist dem Benutzer die Navigation im 3-D-Raum zumutbar?
- Wie hinderlich erscheint das Fehlern der Prozesstopologie?

- Wie ist die Gesamtakzeptanz des neuen Visualisierungssystems?

Bis auf die Darstellung der Vorgeschichte, deren Nutzen von den Probanden nur bedingt erkannt wurde, wurden alle neuen Merkmale des 3-D-Prozessinformationsdisplays positiv bis überdurchschnittlich gut bewertet. Dies trifft insbesondere auf die Darstellung der Dringlichkeitsanzeige zu. Die Navigation im 3-D-Raum wurde von den Probanden als unproblematisch angesehen - dies insbesondere aufgrund der vordefinierten Betrachtungspunkte. Aus Sicht der Probanden werden durch diese Darstellung das Erkennen von Abweichungen und die Notwendigkeit eines Eingriffes stark vereinfacht und beschleunigt. Aufgrund der Verknüpfung zu den traditionellen Prozessgrafiken wird das Fehlen der Prozessstopologie in der 3-D-Darstellung als nicht negativ bewertet. Die positive Bewertung der Fragen zur Akzeptanz und zum Gesamteindruck vom 3-D-Visualisierungssystem rundet das Ergebnis der Befragung ab. Auch hier wurde das System durchgängig als gut bis überdurchschnittlich gut bewertet. Parallel zur Evaluation des 3-D-Visualisierungssystems unter Zuhilfenahme des oben beschriebenen Fragebogens und der Erprobung am realen System wurde das Visualisierungssystem auf verschiedenen Messen, Tagungen und Kongressen vorgestellt. In den sich daraus ergebenden Diskussionen wurden insbesondere die beiden Punkte Prozessstopologie und Kopplung zu den traditionellen Visualisierungsverfahren erörtert. Das Problem der starken Abstraktion des Visualisierungsverfahrens und der daraus folgende Wegfall der Darstellung der Prozessstopologie wurde zwar als schwierig bewertet, die allgemeine Meinung war jedoch, dass der Gewinn zusätzlicher Informationen – wie die Dringlichkeit des Eingriffes, die Geschwindigkeit und Richtung der Abweichung und die schnelle Erfassbarkeit des Prozesszustandes – den Verlust der Prozessstopologie kompensieren. Der Mehrwert dieses Darstellungsverfahrens gegenüber den traditionellen Darstellungsverfahren wurde in den Expertengesprächen im Hinblick auf die Funktion als Übersichtsdarstellung bestätigt. Durch die im Prototypen integrierte Kopplung mit den traditionellen Visualisierungsverfahren wird der Anspruch an eine Übersichtsdarstellung ebenfalls erfüllt. Einzig die Implementierung der 3-D-Welt mittels VRML wurde, wie bereits in Abschnitt 5 diskutiert, aus Performance-Gründen als problematisch angemerkt.

## **7 Zusammenfassung und Ausblick**

Mit den in den Abschnitten 2 und 3 erläuterten Untersuchungen sowie dem in Abschnitt 4 vorgestellten Verfahren konnte gezeigt werden, dass eine ikonbasierte kognitiv-szenische Visualisierung den Benutzer in einer Übersichtsdarstellung auf einfache Art und Weise mit den zur Bewertung des Prozesszustandes notwendigen Informationen versorgen kann. Der Bediener wird unter Stresseinwirkung entlastet, wodurch potenzielle Fehlerquellen minimiert werden. Die kognitiv-szenische Visualisierung soll dabei bestehende Visualisierungsformen nicht ersetzen, sondern ergänzen und als abstrakte und übergeordnete Visualisierungskomponente verstanden werden. Die Auswertung der Ergebnisse der Benutzerbefragung in Abschnitt 6 hat den positiven Einfluss des 3-D-Visualisierungsverfahrens auf das Bedienen komplexer Anlagen bestätigt. Die in Abschnitt 5 angesprochene Umsetzung der 3-D-Welt mittels einer schnelleren 3-D-Engine wie z. B. OpenGL wird die Bedienbarkeit des Systems weiter verbessern. Die Anzahl der vom Bediener problemlos erkennbaren Ikonen, der Einfluss der dreidimensionalen Visualisierung auf die mentale Belastung des Bedieners und die Möglichkeiten sowie Probleme der Navigation im dreidimensionalen Informationsraum sind Gegenstand weiterführender Untersuchungen.

## Literaturverzeichnis

- [1] Alty, J. L.; Elzer, P.; Holst, O.; Johannsen, G.; Savory, S.: Literature und User Survey of Issue Related to Man-Maschine Interfaces for Supervision and Control Systems. Projektbericht Nr. 1. Kassel: GhK-Universität, 1985
- [2] Anderson, J. R.: The adaptive character of thought. 1. Aufl., Hillsdale, NJ: Lawrence Erlbaum, 1990
- [3] Berheide, W.; Dörfel, G.; Döring, B.: Wissensbasierte Benutzerunterstützung in komplexen Mensch-Maschine-Systemen. In: Verlässlichkeit von Mensch-Maschine-Systemen; Hrsg. von Willumeit, H.-P.; Kolrep, H.; Berlin: Technische Universität Berlin, 1995, 61-72
- [4] Beuthel, M. C.: Dreidimensionale Prozessvisualisierung zur Führung technischer Anlagen am Beispiel eines Kohlekraftwerkes. TU Clausthal, Diss. 1997
- [5] Elzer, P. F.: Neuere Entwicklungen auf dem Gebiet der Mensch-Maschine-Schnittstelle. In: at – Automatisierungstechnik, 49 (2001), Heft 4, München: Oldenbourg, 178-186
- [6] Grams, T.: Bedienfehler und ihre Ursachen (Teil 1). In: atp – Automatisierungstechnische Praxis 40 (1998), Heft 3, München: Oldenbourg, 53-56
- [7] Hollnagel, E.; Bye, A.; Hoffmann, N.: Coping with complexity – strategies for information input overload. In: Proceedings of Cognitive Systems Engineering in Process Control (CSEPC) 2000, Taejon, Korea, Nov. 22-25, 2000, 264-268
- [8] Komischke, T.: Informationsüberflutung in der Prozessführung. In: atp – Automatisierungstechnische Praxis, 9 (2003), München: Oldenbourg, 74-79
- [9] Völkel, A.: Prozess- und Aufgabenanalyse in einer Destillationskolonne der Bayer AG Dormagen. Kognitionsbezogene Prozessvisualisierung/Fachgebiet für Systemtechnik und Mensch-Maschine-Systeme. Kassel: GhK-Universität, 1999 (Bericht zum DFG-Forschungsprojekt)
- [10] Wittenberg, C.: Das Konzept der virtuellen Prozessvisualisierung für die industrielle Prozessführung. In: at – Automatisierungstechnik, 4 (2002); München: Oldenbourg, 178-186



## Intelligente Datenanalyse für die optische 3D-Inspektion

Dr.-Ing. Tilo Lilienblum, Dr.-Ing. Stefan Schünemann, Phone: ++49.391-6117-300, Fax: ++49.391-6117-301, E-Mail: 3D@inb-vision.com, INB Vision AG, Leipziger Str. 44, Zenit-Gebäude, D-39120 Magdeburg

**Abstract:** Im Beitrag wird ein optisches 3D-PDM-System der INB Vision AG vorgestellt. Mit diesem System ist es möglich, eine Qualitätskontrolle der Oberflächenform im Produktionsprozess durchzuführen. Die Inspektion ist möglich auf unterschiedlichen Oberflächen (z.B. Metall, Kunststoff), für unterschiedliche Produktionstechnologien (z.B. Guss, Spritzguss, Druckguss, Tiefziehen) und für unterschiedliche Inspektionsaufgaben (z.B. Formkontrolle, Vollständigkeit, Funktion).

Das Inspektionssystem löst die anspruchsvolle Aufgabe mit einer leistungsfähigen Software, die aus zwei Modulen besteht: Ein 3D-Vision-System ist zunächst in der Lage, durch optische Messung 3D-Daten mit einem bezogen auf das Messvolumen sehr geringen Messrauschen zu gewinnen. Ein intelligentes Software-Tool detektiert in den ermittelten 3D-Daten kleine Fehler wie z. B. Beulen, Dellen und Einfallstellen bei geringem Konfigurationsaufwand.

Im Beitrag werden die entwickelten und patentierten Technologien sowie deren Ausführung vorgestellt. Insbesondere wird das Software-Tool zur Auswertung der 3D-Daten beschrieben und die IT-Integration dieser Software erläutert.

**Keywords:** Optische Messtechnik, Automatische Inspektion, 3D, IT-Integration

### 1 Einleitung

Optische 3D-Analyseverfahren (z.B. [1, 2, 7]) haben sich mittlerweile in vielen Bereichen der Qualitätskontrolle etabliert. Ein übliches Verfahren ist die Generierung einer Punktwolke mit einem optischen Sensor und der Vergleich der Daten mit CAD-Daten.

Eine objektive Analyse der Oberflächenform im Produktionsprozess ist jedoch oft schwierig, da z. B. bei Design-Anbauteilen aufgrund von zulässigen Toleranzen Oberflächenformfehler durch einen Vergleich mit einem CAD-Modell nur subjektiv bewertet werden können. Viele solcher Anbauteile, sowohl aus Metall, wie auch aus Kunststoff, erhalten ihre eigentliche Form erst im verbauten Zustand. Es ist deshalb auch oft schwierig, solche Teile auf wenige Mikrometer genau zu justieren und zu prüfen.

Abb. 1 verdeutlicht die Relation zwischen zu detektierendem Fehler und der Toleranz der Oberflächenform.

Es gibt eine Reihe von Verfahren, mit denen im 3D-Datensatz Fehler detektiert werden können. Zunächst kann ein Vergleich mit CAD-Daten vorgenommen werden. Bei großen, nicht starren Teilen werden dabei aber vor allem die Toleranzen des Prüfteils und der Justage sichtbar gemacht. Verfahren, die die lokale Geometrie auswerten [3, 6], versagen an Kanten oder starken Krümmungen der zu prüfenden Teile. Je komplexer das zu prüfende Teil ist, desto mehr Aufwand muss man in diesem Fall in die Intelligenz des Fehlerdetektionsalgorithmus stecken. Üblich ist es, bestimmte Fehler einzulernen. Auf diese Fehler kann dann eine automatische Prüfung erfolgen. Treten neue Fehler auf, können diese zunächst nicht automatisch erkannt werden.

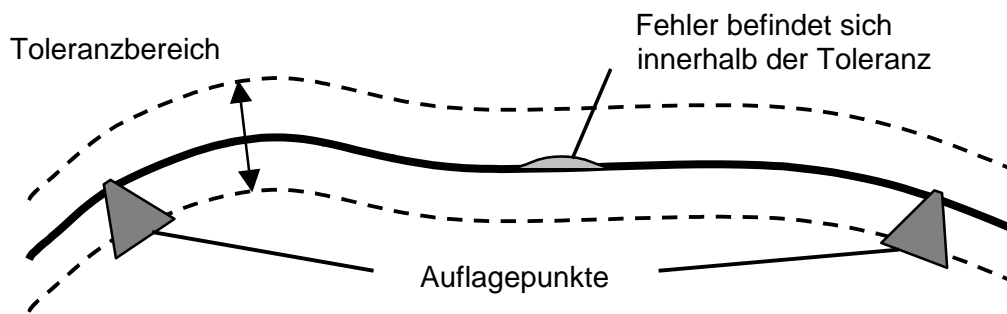


Abb. 1: Beispiel für die Fehlergröße in Relation zur Toleranz des Werkstücks

## 2 Automatische Fehlererkennung mit der Speed Gaging Technologie

Herzstück der Inspektion mit einem 3D-Product Defect Mapping System der INB ist eine intelligente Software-Technologie basierend auf einem Assoziativspeicher, welcher in der Lage ist, in den 3D-Messdaten schnell und sehr effizient Fehler zu detektieren. Dieses Speed Gaging Modul verarbeitet die Daten des optischen INB-Sensors, kann aber auch mit anderen Sensoren betrieben werden (z.B. Infrarot-Sensoren, Schichtdickensensoren, Deflektrometrische Sensoren).

Das *intelligente Softwaretool* der INB zur automatischen Detektion von Fehlern in den 3D-Daten ist ein lernendes System. In einer oder mehreren Trainingsphasen werden Gut-Teile vermessen und das System mit diesen Daten trainiert. Bei der Inspektionsphase im Produktionsprozess kann dann mit dem so konfigurierten System die Vermessung und Auswertung erfolgen.

Abb. 2 verdeutlicht den prinzipiellen Ablauf der Fehlererkennung. Zur Erkennung von lokalen Abweichungen der Oberflächenform eines Prüfobjektes von einer vorgegebenen Form wird folgendermaßen vorgegangen:

In der *Lernphase* (Abb. 2, linke Seite) werden mehrere (fehlerfreie) Meisterstücke in unterschiedlichen Justierungen vermessen. Diese Messdaten werden in einem künstlichen neuronalen Netz [4, 5] gespeichert. Dadurch wird die gesamte Bandbreite der möglichen Variationen und Justierungen der Meisterstücke erfasst.

In der *Recallphase* (Abb. 2, rechte Seite) wird das Prüfobjekt vermessen. Die Messwerte werden in das neuronale Netz eingegeben. In den Gewichten des neuronalen Netzes sind nach der Lernphase Meisterstücke in unterschiedlichen Positionen und globalen Verformungen gespeichert.

Das neuronale Netz [4, 5] ist ein Backpropagation Netz mit 3 Schichten. Schicht 1 ist die Eingabeschicht. Die Z-Koordinaten jedes 3D-Messpunktes bilden die Eingabewerte. In der zweiten Schicht ist die Anzahl der Neuronen stark begrenzt. Über die Anzahl kann die Variabilität des Systems gesteuert werden. In der 3. Schicht erfolgt die Ausgabe der Z-Koordinaten.

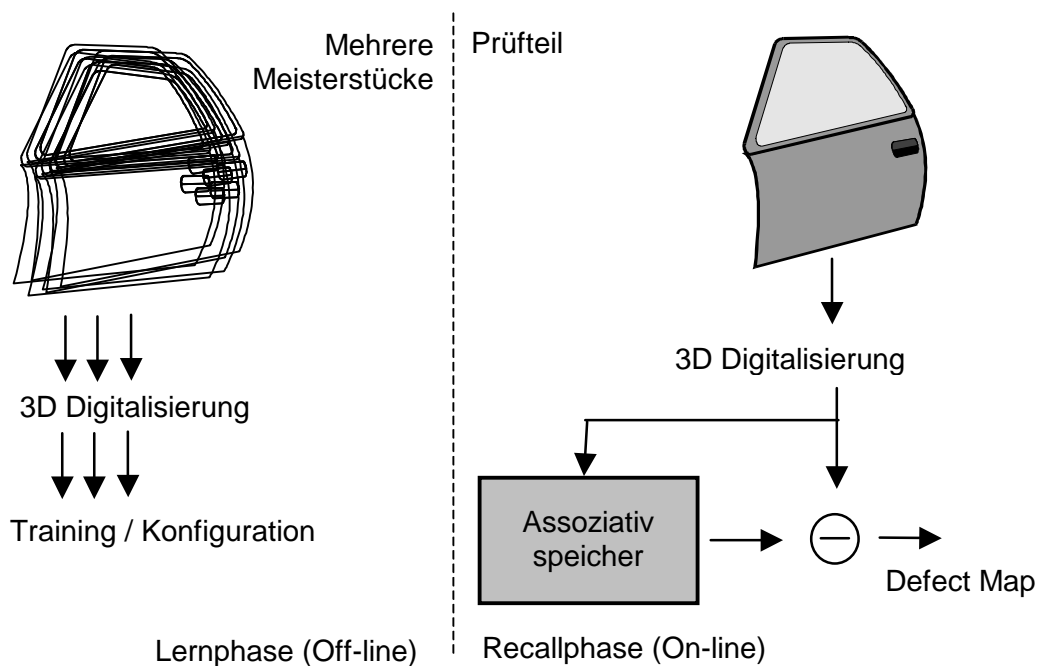


Abb. 2: Automatische Fehlerdetektion mit dem Speed Gaging Module

Am Ausgang des neuronalen Netzes bleiben beim Recall im wesentlichen nur diejenigen Anteile der Messdaten übrig, die den gespeicherten Meisterstücken entsprechen. Fehler, also Abweichungen von diesen Meisterstücken werden geglättet. Nach Differenzbildung zwischen den Daten des gemessenen Werkstücks und den durch das neuronale Netz verarbeiteten Daten bleiben die gesuchten Fehler übrig und können mit einfachen Standard-Verfahren der Bildverarbeitung (z.B. Filterung, Segmentierung) vom Hintergrund getrennt werden.

Mit dieser Methode wird demnach zu jedem vermessenen Prüfobjekt ein zugehöriges Meisterstück generiert, welches in Position, globaler Verformung und Ausführung dem vermessenen Objekt weitgehend gleicht.

Beim Training ist die Variation der Umgebungs- und Produktionsbedingungen (Temperatur, Prozessparameter) zu berücksichtigen. Die für das Training bereitgestellten Teile sollten (wenn vorhanden) aus unterschiedlichen Chargen bereitgestellt werden.

Als Inspektionsergebnis erhält man eine Defect Map des überprüften Bereiches. Die Defect Map zeigt nur die Fehler auf der Oberfläche an. Die Kontur des Werkstücks sowie zulässige Lage- und Formtoleranzen werden unterdrückt.

### 3 Manuelles Inspektionssystem 3D-PDM

Basierend auf der in Punkt 0 vorgestellten Technologie wurde das 3D-PDM (Product Defect Mapping) System entwickelt. Bild 3a und b zeigen zwei Ausführungen des Gerätes. Als manuelles System wurde ein Tischgerät (Messvolumen 210x150x100 mm<sup>2</sup>, Bild 3a) sowie ein modulares System (Messvolumen frei definierbar bis 0.5 m<sup>2</sup>, Bild 3b) aufgebaut. Mit dem modularen System ist eine Anpassung an spezielle Anforderungen möglich. Die

T. Lilienblum

Automatisierung mit Hilfe des 3D-PDM Systems wird in Abschnitt 0 erläutert.

Die Ansteuerung und Kalibrierung des Inspektionssystems sowie die anschließende Auswertung und Präsentation der Inspektionsergebnisse erfolgt mit der Software DefMap3D. Es können die aufgenommenen 3D-Daten sowie das Ergebnis der Inspektion, die Defect Map, in verschiedenen Darstellungen angezeigt werden. Abb. 4 zeigt eine Ansicht von aufgenommenen 3D-Daten und eine Ansicht der Defect Map eines geprüften Teiles.

Die Fehlstellen können zusätzlich durch den im System integrierten Projektor optisch auf dem Werkstück angezeigt werden. Über eine Signalampel kann ein IO- oder NIO-Signal ausgegeben werden.

Bei einigen Einsatzfällen, z.B. bei der Produktion von hochwertigen Teilen im Spritzguss, hat sich auch eine manuelle Überwachung der Defect Map als hilfreich herausgestellt. Das System dient als „Lupe“, also zur Verbesserung der manuellen Qualitätskontrolle.

Falls keine IO-Teile für das Training zur Verfügung stehen, können auch CAD-Daten eingelesen werden. Die Defect Map wird dann durch Vergleich der 3D-Daten gegen CAD gebildet. Das ist zum Beispiel in der Entwicklungsphase von Werkstücken wichtig, da hier noch keine IO-Teile zur Verfügung stehen, aber die Technologie bereits in den Prozess mit eingebunden werden kann. Weiterhin gibt es Prüfaufgaben, bei denen nur ein Vergleich gegen CAD stattfinden soll.

Die Ergebnisse können von der Software auch in komprimierter Form als XML-Dateien weitergegeben werden. Mit dieser Thematik wurde sich im Verbundprojekt „Collaborate“ [8] befasst. Die XML-Daten der Inspektion werden in die Collaborate-Plattform nach der Messung eingespeist. Hier können dann die Defect Map oder die Parameter von Fehlern und Abweichungen in tabellarischer Form dargestellt und archiviert werden. Die Daten sind Teil des digitalen Lebenslaufes des Werkstücks. Kunden oder auch Hersteller können so die Qualitätsdaten eines bestimmten Teiles abfragen.

Das Tool kann besonders effektiv im Qualitätsmanagement eingesetzt werden. Somit können Produktionsprozesse kontrolliert, Nacharbeit vermieden und letztlich Kosten eingespart werden. Das Softwaretool arbeitet wie ein „intelligenter Mitarbeiter“, allerdings nach einem vorab definierbaren Vorgehen. Für die Integration in die Linie ist die Software DefMap3D modular aufgebaut und kann entsprechende Industriekomponenten ansteuern.



Bild 3a: 3D-PDM Desk



Bild 3b: 3D-PDM Modulares System (Sensorkopf)

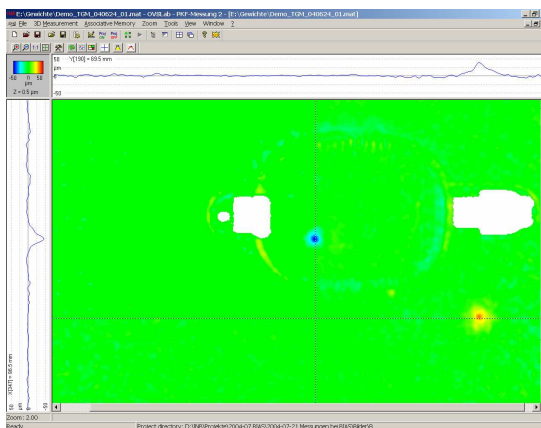
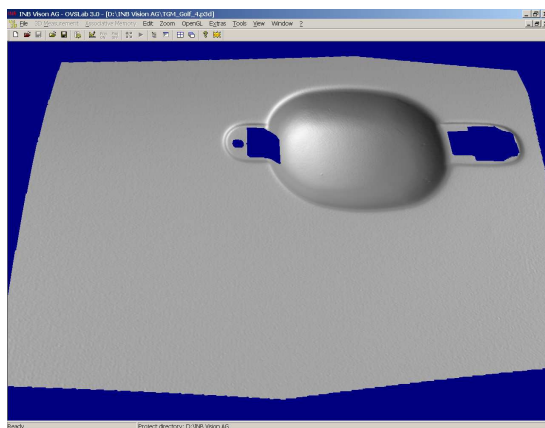


Abb. 4: 3D-Ansicht und Ansicht der Defect Map mit der Software DefMap3D

## 4 Automatisierung und IT-Integration

Es ist nicht erforderlich, dass das Training auf der gleichen Anlage, auf der auch die Inspektion erfolgen soll, durchgeführt wird. Es können auch Trainingsdaten an bestimmten Stellen eingelernt und an anderen Stellen für die Inspektion verwendet werden.

Den Bezugspunkt für die Übertragung der Daten bieten dabei zunächst die Referenzpunkte. Die Ausrichtung des Objektes muss aber nur im Millimeterbereich erfolgen. Unterschiede, die sich durch die Vermessung an unterschiedlichen Orten ergeben können mit antrainiert werden und beeinflussen dann nicht mehr die Qualität der Fehlererkennung.

Der Assoziativspeicher ist deshalb ideal geeignet, um ihn in als zentrales Element in der Kontroll- und Automatisierungsebene einzusetzen.

Abb. 5 zeigt die Einbindung des Assoziativspeichers in eine IT-Infrastruktur. Der Assoziativspeicher bildet das zentrale Element. Er erhält seine Daten aus einer Datenbank, die mit einem zentralen IT-System verbunden ist. Zusätzlich zum INB Sensor können auch

T. Lilienblum

andere optische oder auch nichtoptische Sensoren für die Messung von verschiedensten Messgrößen eingebunden werden.

Eingebunden wird die Technologie in eine Systemsteuerung, die mit dem zentralen IT-System verbunden ist und die Mechanisierung ansteuert.

Mit den Ergebnissen kann der Produktionsprozess optimiert werden. Die Qualität kann unabhängig von der Tagesform eines Auditors verfolgt werden und an das Produktionsmanagement weitergeleitet werden. Ebenso können neue Vorgaben schnell und unkompliziert umgesetzt werden, da nur der zulässige Fehler geändert werden muss (ERP-Ebene).

Auf der Kontroll- und Automatisierungsebene können die Steuerungsparameter geändert werden.

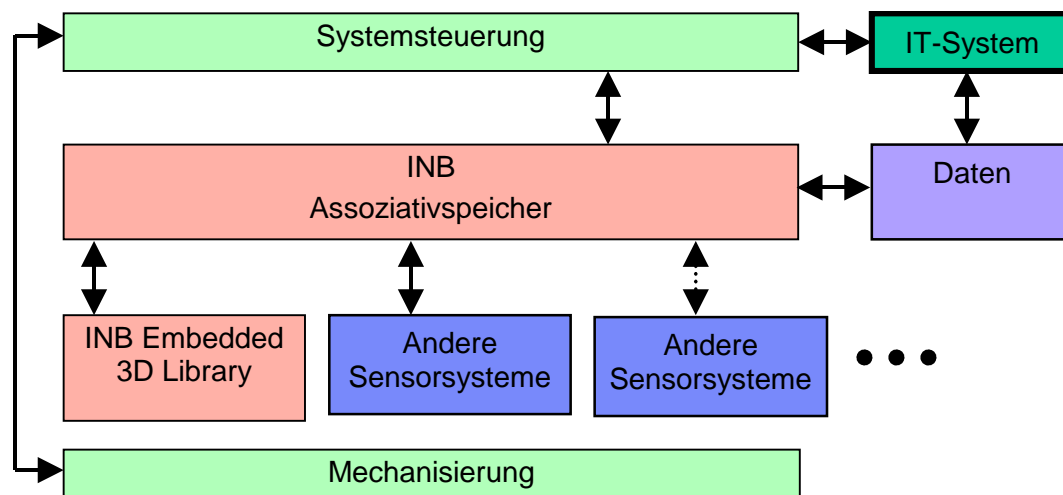


Abb. 5: Der Assoziativspeicher in einer IT- und Automatisierungs-Infrastruktur

Abb. 6 zeigt eine Möglichkeit für die logistische Integration der 3D-PDM-Technologie in die Prozesskette. Als Beispiel wurde die Karosseriefertigung gewählt. Zunächst werden an bestimmten Punkten (am Rohteil oder nach dem Zusammenbau) 3D-Daten erfasst. Diese werden gespeichert. Wichtig ist, dass eine spätere Zuordnung der Daten zu den Teilen möglich ist. Nach der Lackierung und Endkontrolle der Teile erfolgt die Information, ob das entsprechende Teil für die Konfiguration verwendet werden darf. Auf diese Weise ist die Datenbasis immer aktuell und es können sich keine Fehler einschleichen.

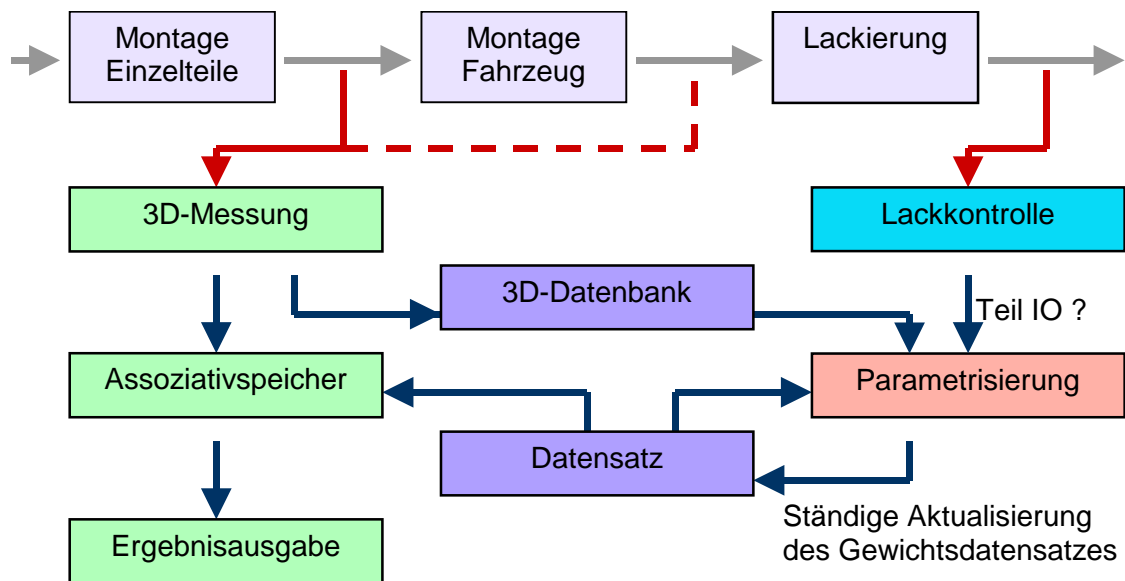


Abb. 6: Logistische Integration in die Prozesskette

## 5 Zusammenfassung

Durch die Entwicklung des optischen 3D-PDM-Systems werden bisherige manuelle Kontrollen, die oftmals durch zusätzliche Oberflächenbehandlung unterstützt werden müssen, überflüssig. Die Inspektion erfolgt bei den Anwendern der 3D-PDM-Technologie auf einer objektiven Basis.

Diese Inspektionssysteme lösen die anspruchsvolle Aufgabe mit einer leistungsfähigen Software, die aus zwei Modulen besteht: Ein 3D-Vision-System ist zunächst in der Lage, durch optische Messung 3D-Daten mit einem bezogen auf das Messvolumen sehr geringen Messrauschen zu gewinnen. Ein intelligentes Software-Tool detektiert in den ermittelten 3D-Daten kleine Fehler wie z. B. Beulen, Dellen und Einfallstellen schnell und aufwandsreduziert.

Eine besondere Leistungsfähigkeit des 3D-PDM-Systems ist, dass die Konfiguration für ein bestimmtes Prüfteil über einen Trainingsprozess erfolgt und für eine große Teilevielfalt konfiguriert werden kann. Das bedeutet, es ist durch den Anwender keine spezielle Programmierung erforderlich. Die Bedienung und Einrichtung des Systems wird nach einer Schulung durch Facharbeiter vor Ort durchgeführt. Bei Bedarf erfolgt der Service über eine Fernwartungsverbindung. Für die Justierung der Teile reichen einfache Anschläge oder mechanische Aufnahmen. Bei einem Teilewechsel müssen nur die Konfigurationsdaten für das neue Teil geladen werden.

Als Inspektionsergebnis erhält man eine Defect Map des überprüften Bereiches. Die Defect Map zeigt nur die Fehler auf der Oberfläche an. Die Kontur des Werkstücks sowie zulässige Lage- und Formtoleranzen werden unterdrückt. Die Fehlstellen können zusätzlich durch den im System integrierten Projektor optisch auf dem Werkstück angezeigt werden. Weiterhin können Kennwerte zu den detektierten Fehlern ausgegeben oder für statistische Auswertungen gespeichert werden. Im einfachsten Fall wird nur ein IO- oder NIO-Signal, z. B. über eine Signallampe, ausgegeben.

In der Produktion können durch das 3D-PDM-System erheblich Kosten eingespart werden.

T. Lilienblum

Bei einem Einsatzfall im Spritzguss kann beispielsweise ein kompletter Arbeitsgang entfallen, da die Qualität bereits vor der Lackierung effizient erfasst werden kann.

Das neue Verfahren wurde für unterschiedliche Messvolumen (Messfläche pro Einzelsystem: 100 cm<sup>2</sup> ... 0.5 m<sup>2</sup>) und Materialien (Aluminium- oder verzinktes Stahlblech, Aluminiumguss oder verschiedene Kunststoffe) in der Industrie validiert.

Die 3D-PDM Systeme können in verschiedenen Varianten ausgeführt werden. So stehen als manuelle Systeme z. B. ein Tischgerät (Messvolumen 210x150x100 mm<sup>2</sup>) sowie ein modulares System (Messvolumen frei definierbar bis 0.5 m<sup>2</sup>) zur Verfügung. Für die Integration in die Linie ist die Software DefMap3D modular aufgebaut und kann entsprechende Industriekomponenten ansteuern.

## Literatur

- [1] Albrecht, P.; Michaelis, B.: Stereo Photogrammetry with Improved Spatial Resolution. The 14th International Conference on Pattern Recognition (ICPR '98), Brisbane, Queensland, Australia, 17.08-20.08.1998
- [2] Albrecht, P.; Michaelis, B.: Verfahren zur dreidimensionalen optischen Vermessung von Objektoberflächen. Patent: DE 19928341.
- [3] Karbacher, St.; Babst, J; Häusler, G.; Laboureux, X.: Visualization and Detection of Small Defects on Car-Bodies. In: B. Girod, H. Nieman, H.-P. Seidel, Hrsg., Vision, Modelling and Visualization ,99, Infix Verlag, Sankt Augustin, 1999, S. 1-8
- [4] Lilienblum, T.; Michaelis, B.; Günther, G.; Bauer, G.; Küchel, M.: Verfahren zur Erkennung von Abweichungen der Oberflächenform ... Patent: DE 19753620.
- [5] Lilienblum, T.: Aufbereitung von 3D-Meßwerten unter Nutzung künstlicher neuronaler Netze. Dissertation. Shaker Verlag, 1999
- [6] Recknagel, R.-J.: Defekterkennung an Oberflächen mittels Waveletmethoden. Hochschulschrift: Jena, Univ., Diss
- [7] Wahl, F.: A Coded Light Approach for Depth Map Aquisition. 8. DAGM-Symposium Paderborn 1986, Springer Verlag 1986
- [8] <https://www.projekt.collaborate-online.de>

*Unser Dank gilt dem BMBF, das Teile der hier dargestellten Arbeiten im Rahmen der INNO-Regio-Initiative gefördert hat (FKZ 0311233F).*



# Lösungsansatz für eine ereignisgesteuerte Prozessdiagnose

Jürgen Jörn, Phone: ++49.3943-659-321, Fax: ++49.3943-659-399, E-Mail: jjoern@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

**Abstract:** Es wird ein Wissensansatz gezeigt, wie auf der Basis eines Modells einer technischen Anlage ein Anlagenüberwachungssystem entwickelt wird. Schwerpunkt ist die besondere Art der Konsistenzprüfung.

Unter Zuhilfenahme von Teilen eines Erreichbarkeitsgraphen, welcher am Systemmodell aufgestellt wird, kann eine Ablauffolge generiert werden. Der Programmalgorithmus eines Zustandsanalysetools ermöglicht, aus den Modellzustandsübergängen identifizierbare unterscheidbare Zustandsänderungen zu gewinnen. Mit dem aufgestellten Beobachter wird ein Überwachungsansatz gebildet und eine Zustandsvergleichsmethode entwickelt. Durch den Vergleich kommt es zur Bildung einer Fehlermatrix mit Fehlerlokalisationszeigern. In der anschließenden Diagnoseauswertung werden diese durch ein Fehlerreaktionsprogramm analysiert.

**Stichworte:** Modell, Beobachter, Zustandsvergleich, Diagnose

## 1 Einleitung

In der Automatisierungstechnik ist eines der wichtigsten zentralen Themen die Überwachung industrieller fertigungstechnischer Anlagen auf einwandfreie Funktion. Funktionsstörungen haben zum Teil gravierende Auswirkungen, die von Produktionsunterbrechungen über Schäden am Produkt oder der Anlage bis zur Gefährdung von Menschenleben reichen. In vielen Fällen können durch rechtzeitiges Erkennen die negativen Auswirkungen einer Störung gemildert oder sogar verhindert werden. Systeme, die die Anlagen bzw. Anlagenteile überwachen, werden als Diagnose- oder Anlagenüberwachungssysteme bezeichnet [Lu95].

Der grundlegende Gedanke, der es erlaubt, technische Anlagenüberwachungssysteme auf Basis von Anlagenmodellen zu realisieren, ist, eine Abweichung als Inkonsistenz zwischen Anlagenmodell und Anlagenzustand zu definieren. Während des Betriebs der Anlage wird der Ablauf in Schrittfolgen zerlegt. Die aktuellen Zustände der Anlage werden in eine Matrixnotation umgewandelt, um diese so mit der Matrixschreibweise des Modells auf Gleichheit zu untersuchen. Dieser neue Ansatz zur Prüfung der Inkonsistenz erweist sich als sehr brauchbar. Durch die Verwendung von mathematische Beziehungen ist auch eine Algorithmierbarkeit gegeben. Ergibt ein Vergleich eine Inkonsistenz, so ist eine Störung bzw. ein Fehler in der Anlage aufgetreten.

Unter Zuhilfenahme von Steuerzuständen der projektierten Steuereinrichtung werden bestimmte innere Zustände eines ereignisdiskreten Systems abgebildet. So ist es in bestimmten Grenzen möglich, System- bzw. Anlagenzustände mit Modellzuständen zu vergleichen. Werden das Anlagenmodell und der Anlagenzustand korrekt formuliert, so kann man davon ausgehen, dass beide, jeweils für sich betrachtet, widerspruchsfrei sind. Bei der Vereinigung von Modellzustand und Anlagenzustand kann daher eine Inkonsistenz nur auftreten, wenn der aktuelle Anlagenzustand sich im Widerspruch zum Anlagenmodell befindet. Um ein im obigen Ansatz beruhendes Überwachungssystem zu realisieren, muss während des Anlagenbetriebes in jeder Schrittfolge der Vergleich auf Konsistenz durchgeführt werden. Aber auch eine spätere Auswertung und Fehlerermittlung ist durchführbar.

## 2 Lösungsansatz mittels modellbasierter Systemüberwachung

### 2.1 Gegenstand der zu untersuchenden Thematik

Gegeben ist ein ereignisdiskretes System. Die kausalen Veränderungen der Stellgrößen werden in einem geschlossenen Wirkungskreis durch die projektierte Steuerung berechnet bzw. festgelegt. Durch diese Festlegung entsteht eine Ereignisfolge bei der Bildung der inneren Zustände und der Stellgrößen. Die gebildete Ereignisfolge ignoriert die Werte der Zeitpunkte und stellt lediglich dar, in welcher Reihenfolge die diskreten Zustände und Stellgrößen erreicht werden. Die gebildete Reihenfolge  $k$  mit  $k \in N(1,2,3,\dots)$  wird mit einem Zähler durchnummeriert. Der Wert  $k$  beschreibt nicht den exakten Zeitpunkt eines Signalwertwechsels, sondern nur die Reihenfolge dieser Wechsel. Das Verhalten äußert sich in Folgen diskreter Schaltvorgänge und wird als Ereignisfolge bezeichnet [Lu95].

Die Gesamtheit aller inneren Zustände des ereignisdiskreten Systems ist nicht beobachtbar. Beobachtbar ist die Folge der Zustandsänderung der projektierten Steuerung  $z_{ps}(k)$  (sie wird hier auch als Folge der Steuerungszwischenzustände bezeichnet) und die Wertefolge aller Stellgrößen.

Es soll eine „Beobachtbarkeit“ entwickelt werden, die in einem Beobachter unter Verwendung der jeweiligen Eingangs- und Ausgangssignale, der Steuerungszwischenzustände und den bereits ermittelten angenommenen inneren Zuständen im existierenden ereignisdiskreten Systems den realen Steuerungsablauf im System abbildet. Der Beobachter rekonstruiert also aus dem Verlauf der Eingangs-, Ausgangsgrößen und Steuerungszwischenzustände über einen angegebenen Zeithorizont bestimmte innere Zustände des ereignisdiskreten Systems. Alle hieraus ermittelten Zustände  $z_B(k)$  werden in diesem Beobachter nach dem jeweiligen Auftreten der Ereignisse aufgelistet und zusammengefasst. Es gibt keine Rückwirkung vom Beobachter auf das System.

In einem zweiten Schritt vergleicht ein verifiziertes Systemmodell in der Ereignisfolge seine Parameter  $u(t)$ ,  $z_M(k)$  und  $w(k)$  mit denen des Beobachters und bestimmt die jeweilige Abweichung daraus. Bei jedem neuen Zählerwert  $k$  wird das auftretende Ereignis (Eingangs-, Zustands- und Ausgangsereignis) miteinander verglichen und in eine Matrixnotation abgelegt. Treten Abweichungen auf, erfolgt im Anschluss eine detaillierte Fehleridentifikation mit Hilfe von Diagnosealgorithmen, die den möglichen Fehler bzw. die Fehlerklasse ermittelt [Jö06].

Je nach der Größe des ereignisdiskreten Systems erfolgt der Vergleich in untergeordneten Teilsystemen oder Modulen. Für diese hierarchische Modellierung wird ein Netz-Condition/Event-Systemmodell (NCES-Modell) eingesetzt [Th02].

Über Conditioneingangs- bzw. -ausgangssignale (cin/cout) und über Eventeingangs- bzw. -ausgangssignale (ein/eout) erfolgt die Kopplung der projektierten Steuerung an die Steuerstrecke und an das Modell bzw. den Beobachter.

Abb. 1 zeigt diese Struktur für den Zustandsvergleich.

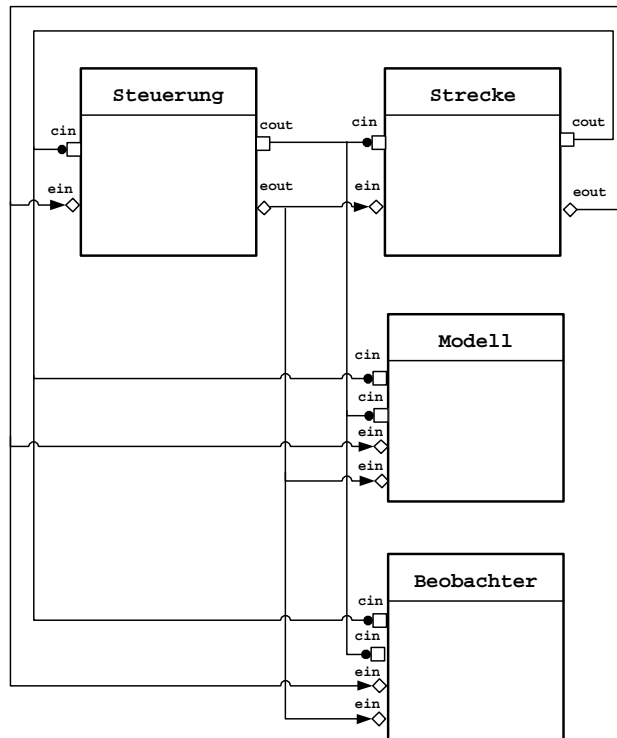


Abb. 1: Struktur für den Zustandsvergleich

## 2.2 Stand des Wissens

In der Literatur gibt es für die Fehlerdiagnose von ereignisdiskreten Systemen, die aus beobachtbaren und zum Teil nichtbeobachtbaren Zuständen und Zustandsübergängen eines Systems eine Prozessdiagnose gewinnen, Veröffentlichungen mit unterschiedlichen Beschreibungsmöglichkeiten.

In [SaSeLaSi95, Sa95] sind z. B. aufgeführt:

- Fehlerbaumdiagnosen,
- Lösungsmethoden, die auf analytischer Redundanz basieren,
- DES-Theorie nach Sampath, Lafortune und Teneketzi,
- Modellbasierte Lösungsmethoden und
- Expertensysteme.

Eine Fehlerdiagnose mit der DES-Theorie nach Sampath, Lafortune und Teneketzi [SaLaTe96] ist für spontane und leichte Fehler geeignet. Sie ist nicht geeignet zur Diagnose von Fehlern die zum Komplettausfall des gesamten Systems bzw. des Prozesses führen. Diese Diagnostiktheorie basiert auf formalen Sprachen unter Anwendung von endlichen Zustandsmaschinen.

Sie kann sowohl offline als auch online diskrete Ereignissysteme (DES) diagnostizieren und ist besonders geeignet für Fehler die im Prozessverlauf selbst auftreten.

Ein ähnlicher Ansatz ist in [SaSeLaSi95] beschrieben. Hier wird ein nichtdeterministischer Automat als Beobachter verwendet, der eine Vorhersage des aktuellen Zustandes des zu betrachtenden Systems angibt. In dieser vorgestellten Methode wird der Beobachter außerdem verwendet, um notwendige und ausreichende Bedingungen für eine Diagnostizierbarkeit von Ereignisabläufen festzustellen.

Für die in diesem Beitrag verwendeten NCEs-Strukturen sind Matrixnotationen für eine Fehlerdiagnose noch nicht angewendet worden.

### 2.3 Beobachterstruktur

Für die Entwicklung der Beobachterstruktur werden die jeweiligen Eingangs- und Ausgangssignale und die Steuerungszwischenzustände der aufgestellten Steuerung des existierenden kausalen ereignisdiskreten Systems verwendet. Die daraus ermittelte Zustandsablauffolge soll möglichst genau den realen Steuerungsablauf des Systems abbilden. Zur weiteren Auswertung gelangen die bekannten Zustandsänderungen der Steuerung und alle auswertbaren Ein- und Ausgangsmessglieder (z. B. Endschalter oder Initiatoren) zum Einsatz. Die Elemente der Beobachterablauffolge werden wie folgt definiert:

$$MB = (s_B^{jm}) \quad \forall j \in \{1, 2, \dots, l\} \quad \text{und} \quad \forall m \in \{1, 2, \dots, k\}.$$

Index  $k$  enthält die Prozessschritte, Index  $l$  enthält die Anzahl der verwendeten Betrachtungseinheiten bzw. Teilmodule.

Die Beobachterablauffolge wird durch folgende Schreibweise festgelegt:

$$MB = \begin{pmatrix} s_B^{10} & s_B^{11} & \dots & s_B^{1m} & \dots & s_B^{1k} \\ s_B^{20} & s_B^{21} & \dots & s_B^{2m} & \dots & s_B^{2k} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ s_B^{l0} & s_B^{l1} & \dots & s_B^{lm} & \dots & s_B^{lk} \end{pmatrix}$$

Hier gilt für  $s_B^{12}$  die  $i$ -te Stelle im 1. Teilmodul des Beobachtermodells, welche im 2. Schritt abgefragt wird,  $s_B^{1k}$  ist dann die  $i$ -te Stelle des 1. Teilmoduls im  $k$ -ten Schritt. Die Bezeichnung  $s_B^{lk}$  symbolisiert die  $i$ -te Stelle des  $l$ -ten Teilmoduls im  $k$ -ten Schritt.

### 2.4 Modellablauffolge

Das verifizierte Modell wird durch die Modellablauffolge  $MM$  beschrieben. Für sie gilt ein analoger Aufbau zur Beobachterablauffolge  $MB$ .

Die Elemente der Modellablauffolge werden wie folgt definiert:

$$MM = (s_M^{jm}) \quad \forall j \in \{1, 2, \dots, l\} \quad \text{und} \quad \forall m \in \{1, 2, \dots, k\}.$$

Index  $k$  enthält die Prozessschritte, Index  $l$  enthält die Anzahl der verwendeten Betrachtungseinheiten bzw. Teilmodule.

Die Modellablauffolge wird durch folgende Schreibweise festgelegt:

$$MM = \begin{pmatrix} s_M^{10} & s_M^{11} & \dots & s_M^{1m} & \dots & s_M^{1k} \\ s_M^{20} & s_M^{21} & \dots & s_M^{2m} & \dots & s_M^{2k} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ s_M^{l0} & s_M^{l1} & \dots & s_M^{lm} & \dots & s_M^{lk} \end{pmatrix}$$

Die Bezeichnung  $s_M^{11}$  markiert die i-te Stelle des 1.verwendeten Teilmoduls (bzw. Teilsystems) des Modells im 1. Schritt,  $s_M^{22}$  bezeichnet die i-te Stelle im 2.Teilmodul des 2. Schrittes. Analog zur Beobachterfolge wird mit  $k$  die Anzahl der Schritte und mit  $l$  die Anzahl der verwendeten Teilmodule (bzw. Teilprozesse) angesehen.

## 2.5 Zustandsvergleich

Aus dem Vergleich der verifizierten Modellablauffolge mit der Beobachterablauffolge ergibt sich die Zustandsvergleichsablauffolge **MZV**.

Bei diesem Vergleich werden die Inhalte der einzelnen Elemente  $s_M^{jm}$  der Modellablauffolge mit den Inhalten der jeweiligen Elemente  $s_B^{jm}$  der Beobachterablauffolge gegenübergestellt. Das Ergebnis des einzelnen Vergleiches steht als Inhalt im entsprechenden Element  $s_{ZV}^{jm}$  der Zustandsvergleichsablauffolge, es gilt:

$$\mathbf{MZV} = (s_M^{jm} \text{ XOR } s_B^{jm}) = (s_{ZV}^{jm}) \quad \forall j \in \{1, 2, \dots, l\} \text{ und } \forall m \in \{1, 2, \dots, k\}.$$

1. Fall:  $s_{ZV}^{jm} = 0 \rightarrow$  Gleichheit der Elemente.
2. Fall:  $s_{ZV}^{jm} \neq 0 \rightarrow$  die Inhalte der Elemente sind verschieden, der Inhalt von  $s_{ZV}^{jm}$  wird auf eins gesetzt.

$$\begin{pmatrix} s_{ZV}^{10} & s_{ZV}^{11} & \dots & s_{ZV}^{jm} & \dots & s_{ZV}^{1k} \\ s_{ZV}^{20} & s_{ZV}^{21} & \dots & s_{ZV}^{jm} & \dots & s_{ZV}^{2k} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ s_{ZV}^{l0} & s_{ZV}^{l1} & \dots & s_{ZV}^{jm} & \dots & s_{ZV}^{lk} \end{pmatrix} = \begin{pmatrix} s_M^{10} & s_M^{11} & \dots & s_M^{jm} & \dots & s_M^{1k} \\ s_M^{20} & s_M^{21} & \dots & s_M^{jm} & \dots & s_M^{2k} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ s_M^{l0} & s_M^{l1} & \dots & s_M^{jm} & \dots & s_M^{lk} \end{pmatrix} \text{ XOR } \begin{pmatrix} s_B^{10} & s_B^{11} & \dots & s_B^{jm} & \dots & s_B^{1k} \\ s_B^{20} & s_B^{21} & \dots & s_B^{jm} & \dots & s_B^{2k} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ s_B^{l0} & s_B^{l1} & \dots & s_B^{jm} & \dots & s_B^{lk} \end{pmatrix}$$

## 2.6 Fehlerlokalisationszeiger und Fehlermatrix

Die gebildete Zustandsvergleichsmatrix enthält in der Spalte den durch den jeweiligen Schritt erreichten Vergleich, er wird als Fehlerlokalisierungszeiger **FLZ** definiert. Die Zeilen enthalten die Wertigkeiten des Vergleiches der Zustände von Teilbeobachter und Teilmodell. Die Summe dieser FLZ-Vektoren bilden den beobachteten Fehlervektor oder die Fehlermatrix. Diese Fehlermatrix **FM** ergibt sich in der Größe aus der Zustandsvergleichsablauffolge **MZV**. Die Dimension dieser Matrix wird durch die Anzahl der verwendeten Teilmodule (Zeilenanzahl) und den festgelegten Ablaufschritten (Spaltenanzahl) bestimmt.

Für die Fehlermatrix gilt folgende Definition:

$$\mathbf{FM} = \mathbf{FLZ}_0 + \mathbf{FLZ}_1 + \mathbf{FLZ}_2 + \mathbf{FLZ}_3 + \dots + \mathbf{FLZ}_k \quad \text{mit } k = \text{Anzahl der Schritte.}$$

Ist der Vergleich aus Beobachter und Systemmodell identisch, so ist der Fehler-Vektor in jedem Abfragezeitpunkt ein Null-Vektor. Tritt eine Abweichung, ein Fehlerereignis auf, ergeben sich je nach Fehlerart ganz genau definierte Vektorwerte, die in den einzelnen FLZ des jeweiligen Abfrageschrittes widergespiegelt werden. Der für jeden Schritt und damit für den zugeordneten Zustand entstandene Fehlerlokalisierungszeiger zeigt in seinem Ablauf (  $k = 0, 1, 2, 3, \dots$  ) die Übereinstimmung bzw. die Abweichung des Beobachters zum verifizierten Modell an. Aus den

so ermittelten FLZ-Vektorwerten lassen sich relativ einfach mögliche aufgetretene Fehler erkennen und lokalisieren. Durch die Zuordnung in Schritte und die Aufteilung des Anlagensystems in Teilanlagen und Module ist die Voraussetzung dazu gegeben.

## 2.7 Analyse und Fehlerauswertung

Mit einem NCES-Editor der Uni Halle, [Th02] wird das Modell erzeugt und mit dem Analysewerkzeug SESA der TU Berlin [StaRo02], wird der Erreichbarkeitsgraph des Modells aufgestellt. Im entwickelten Zustands-Analyse-Tool werden danach die Modellparameter mit den Parameter des abgeleiteten Beobachters verglichen und ausgewertet. Der neben dem Modell aufgestellte Beobachter gestattet im Offline-Betrieb im Zustands-Analyse-Tool zusätzliche Simulationsmöglichkeiten.

Ein weiteres aufgestelltes Fehlerreaktionsprogramm ergänzt die Auswertung, die gewonnenen Fehlerlokalisationszeiger werden hier in einer vorher vereinbarten und festgelegten Fehlerliste zugeordnet.

## 3 Anwendung

In diesem Abschnitt soll der aufgestellte Lösungsansatz an einer einfachen Anlage durchgeführt werden. Diese Beispielanlage besteht aus einem Werkstückmagazin und einem Ausschieber (Abb. 2). Der Ausschieber transportiert ein Werkstück (WST) aus dem Magazin (Position A) zu einer Ablage (Position B). Über das Ventil 1Y1 gelangt Druckluft an den Ausschieber. Es sind folgende Aktoren, Sensoren und Messgrößen verwendet worden:

- 1Y1** pneumatischer Ausschiebezylinder mit Federrückstellung
- 1B1** Initiator, Ausschiebezylinder ist in Ausgangsstellung (hinten)
- 1B2** Initiator, Ausschiebezylinder ist ausgefahren (vorn)
- 1S1** WST an Position B angekommen
- 1B3** Initiator, Werkstückmagazin ist leer
- 1S101** Startschalter für den Vereinzlungszyklus

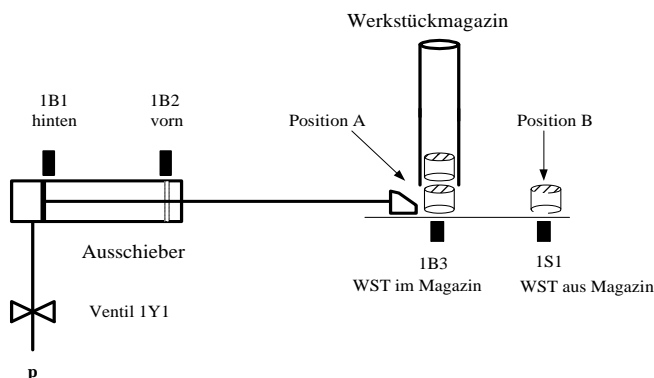


Abb. 2: Beispielanlage „Werkstückmagazin und Ausschieber“

Es werden unterschiedliche Fehlerarten simuliert, 13 unterschiedliche Fehler, die in 7 Fehlergruppen zusammengefasst sind. Für jeden einzelnen dieser simulierten Fehler ist im Anschluss der dazugehörige FLZ-Vektor und die Fehlermatrix ermittelt worden.

### 3.1 Simulierte Fehler:

- F01** - vor dem Start keine Druckluft
- F2a** - nach dem Start keine Druckluft
- F2b** - in der Ausschieberzwischenstellung keine Druckluft, Werkstück hat nur einen Teil der Position aus dem Magazin erreicht, beliebige Stellung zwischen den Endlagen
- F3a** - Ausfall Ventil 1Y1, beim Vorfahren des AS-Zylinders
- F3b** - Ausfall Ventil 1Y1, beim Zurückfahren des AS-Zylinders
- F4a** - Initiator 1B1 defekt, ständig 0-Signal
- F4b** - Initiator 1B1 defekt, ständig 1-Signal
- F5a** - Initiator 1B2 defekt, ständig 0-Signal
- F5b** - Initiator 1B2 defekt, ständig 1-Signal
- F6a** - Endschalter 1S1 defekt, ständig 0-Signal
- F6b** - Endschalter 1S1 defekt, ständig 1-Signal
- F7a** - Initiator 1B3 defekt, ständig 0-Signal
- F7b** - Initiator 1B3 defekt, ständig 1-Signal

### 3.2 Ermittlung der Fehlerlokalisationszeiger und der Fehlermatrix für die simulierten Fehlerarten (Auszug)

Für die entsprechenden Fehler werden aus Übersichtsgründen die einzelnen FLZ-Werte zusammen abgebildet (z. B. 2. und 3. Prozessschritt zusammen). Alle simulierten Fehler (F01 bis F7b) sind unterscheidbar.

#### Fehler F01 – vor dem Start keine Druckluft

(Es ist aus bestimmten Gründen keine Auswertung des vorhandenen Druckluftes möglich.), Fehlerbild: AS-Zylinder bewegt sich nicht nach vorn.

$FLZ_{F01}$ : Teilmodul:

Schritt:  
2. 3.

- Startmodul S101
- ST AS-Zylinder
- Ventil 1Y1
- Ausschiebezylinder
- Messeinrichtungsmodul 1B2
- Messeinrichtungsmodul 1B1
- Messeinrichtungsmodul 1S1
- Messeinrichtungsmodul 1B3
- Modul Position Werkstück

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$FM_{F01} =$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

#### Fehler F03a – Ausfall Ventil 1Y1, beim Vorfahren des AS-Zylinders

Fehlerbild: Kein Vorfahren des AS-Zylinders bei einem elektrischen Ausfall des Ventils 1Y1, oder bei Verklemmung des AS-Zylinders in der Ausgangslage.

$FLZ_{F03a}$ : Teilmodul:

Schritt:  
2. 3. 4.

Startmodul S101	0	0	1
STAS-Zylinder	0	0	1
Ventil 1Y1	0	0	0
Ausschiebezylinder	1	1	1
Messeinrichtungsmodul 1B2	0	1	1
Messeinrichtungsmodul 1B1	1	1	1
Messeinrichtungsmodul 1S1	0	1	1
Messeinrichtungsmodul 1B3	0	1	1
Modul Position Werkstück	1	0	1

$$FM_{F03a} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

## 4 Zusammenfassung

In diesem Beitrag wird ein integriertes Gesamtkonzept zur Modellierung, Analyse, Beobachtung und Diagnose ereignisgesteuerter Systeme aufgestellt. Es zielt darauf ab, Modellbeschreibungen mit NCES durch eine modellbasierte Systemüberwachung in eine ereignisgesteuerte Prozessdiagnose überzuführen.

Durch die Entwicklung eines Beobachters, der unter Zuhilfenahme von aufgestellten Steuerzuständen der eingesetzten Steuereinrichtung bestimmte innere Zustände eines zu untersuchenden ereignisdiskreten Systems abbilden kann, ist die Möglichkeit gegeben, in bestimmten Grenzen System- und Anlagenzustände mit Modellzuständen zu vergleichen.

Da nur universitäre Software eingesetzt wurde, kann eine Aussage über einen komplexen Einsatz z. Z. nur in einem begrenzten Umfang gegeben werden. Untersuchungen bis zu einer Menge von 43 hierarchisch strukturiert angeordneter Module sind erfolgreich durchgeführt worden. Weitergehende Betrachtungen mit einem Schwerpunkt über nichtdeterministisches Verhalten sind als Fortführung dieser Ausarbeitungen unumgänglich.

## 5 Literaturnachweis

- [Jö06] Jörn, J.: *Ereignisgesteuerte Prozessdiagnose durch eine modellbasierte Systemüberwachung*, EKA 2006, Technische Universität Braunschweig, Braunschweig, 29.-31.05.2006.
- [Lu 95] Lunze J.: *Künstliche Intelligenz für Ingenieure*, Band 2: Technische Anwendungen, Oldenbourg Verlag, München 1995.
- [SaLaTe96] Sampath, M. ; Larfortune, S. ; Teneketzis, D. : *A Language-based Approach to Failure Diagnosis of Discrete Event*, IEEE, WODES 96, Edinburgh, UK.
- [Sa95] Sampath, M.: *A Discrete Event Systems Approach to Failure Diagnosis*; PhD thesis, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, 1995.
- [SaSeLaSi95] Sampath, M.; Sengupta, R.; Larfortune, S.; Sinnamohideen, K.; Teneketzis, D. *Diagnosability of discrete event systems*, IEEE Trans. Automatic Control, 40(9):1555-1575, September 1995.
- [StaRo02] Starke; P., Roch, S.: *Analysing Signal-Net Systems*, Informatikbericht 162, Institut für Informatik, Humboldt-Universität Berlin, September 2002.
- [Th02] Thieme, J.: *Symbolische Erreichbarkeitsanalyse und automatische Implementierung strukturierter, zeitbewerteter Steuerungsmodelle*, Dissertation, Martin-Luther-Universität Halle-Wittenberg, 2002.



# **Automatisierungstechnik II**



# Die umweltgerechte Gestaltung von Motoren und deren Bewertung

Prof. Dr. Andrea Heilmann, Phone: ++49.3943-659-312, Fax: ++49.3943-659-399, E-Mail: aheilmann@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

Ute Urban, Phone: ++49.3943-659-360, Hochschule Harz, uurban@hs-harz.de

## Abstract

Electric motor driven systems use 69% of electricity in Europe (EU-15, 1998). Different studies on the consumption characterisation of motor driven systems have shown large potentials for electricity savings up to 181 billion kWh/year in Europe (EU 15). This saving potential can be reached by using high efficiency motors, installing variable speed drives and optimising the application part of the driven system. Nevertheless, the implementation of these measures has been limited by several factors, including higher purchase costs and the lack of knowledge about the economical and ecological benefits.

In order to inform the users in more detail, a new EU-directive demands environmental product declarations.

This paper describes two opportunities to improve energy efficiency in motor driven systems and their ecological assessment by using different software tools.

**Keywords:** motor driven systems, energy efficiency, life cycle assessment, environmental product declaration

## 1 Einleitung

Motorenangetriebene Systeme verbrauchen ca. 65% der elektrischen Energie in Europa. Durch die Verbesserung der Effizienz dieser Systeme sowohl bei der Produktentwicklung als auch bei der Optimierung bestehender Systeme durch unterschiedliche Maßnahmen werden Einsparungen von bis zu 200 TWh (EU-25) pro Jahr erwartet [1]. Energieeinsparungen sind auf der einen Seite die kostengünstigste Art, die Versorgungssicherheit zu erhöhen und die Abhängigkeit von Einfuhren zu verringern. Auf der anderen Seite werden dadurch auch ökologische Folgen reduziert, insbesondere die Entstehung von Treibhausgasen [2].

Diese Überlegungen führten zur Verabschiedung der EuP-Richtlinie (Energy Using Products) [2], die einen europaweit einheitlichen Rahmen für die Festlegung von Anforderungen an die umweltgerechte Gestaltung energiegetriebener Produkte schafft. Wesentliche Ziele der Richtlinie sind:

- die Verringerung des Energieverbrauches durch Festlegung allgemeiner (Anhang I der Richtlinie) und spezifischer (Anhang II der Richtlinie) Ökodesign- Anforderungen
- die verbesserte Information der Verbraucher über Umweltaspekte und Eigenschaften um die Nachfrage.

Die zu berücksichtigenden Maßnahmen sollen sich dabei auf den gesamten, in der Abbildung 1 dargestellten Produktlebenszyklus beziehen, d.h. bei der Rohstoffauswahl und Gestaltung der Motoren sind bereits der Verbrauch in der Nutzungsphase und die Möglichkeiten des Recyclings zu berücksichtigen. Während der Nutzer seine wirtschaftliche Entscheidung aufgrund der Lebenszykluskosten treffen wird, sollen die ökologischen Produktauswirkungen durch die Methode der Ökobilanzierung bewertet werden.

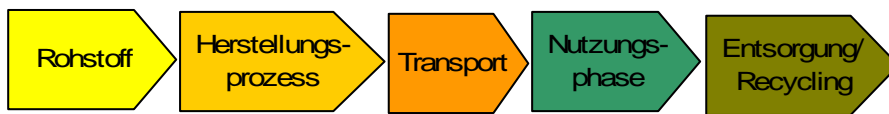


Abb. 1: Darstellung des Produktlebenszyklus [4]

Nachfolgend werden verschiedene Möglichkeiten der Effizienzsteigerung von Motoren vorgestellt und hinsichtlich ihrer ökologischen Auswirkungen verglichen.

## 2 Gestaltung von Motoren und Systemen

Um den Energieverbrauch in der Nutzungsphase zu reduzieren, können prinzipiell drei unterschiedliche Ansätze verfolgt werden [1]. Zum einen sollen die Motoren selbst durch verbesserte Wirkungsgrade effizienter werden. In der Abbildung 2 sind die drei Wirkungsgradklassen EFF3, EFF2 und EFF1 dargestellt, die die Grundlage der Selbstverpflichtung der Motorenhersteller bilden [3]. Es wird deutlich, dass insbesondere bei geringen Nennleistungen Verbesserungen des Wirkungsgrades durch den Übergang von EFF3 zu EFF2 zu erreichen sind. Die Verbesserungen des Wirkungsgrades beim Übergang von EFF2 zu EFF1 sind deutlich geringer [3,4].

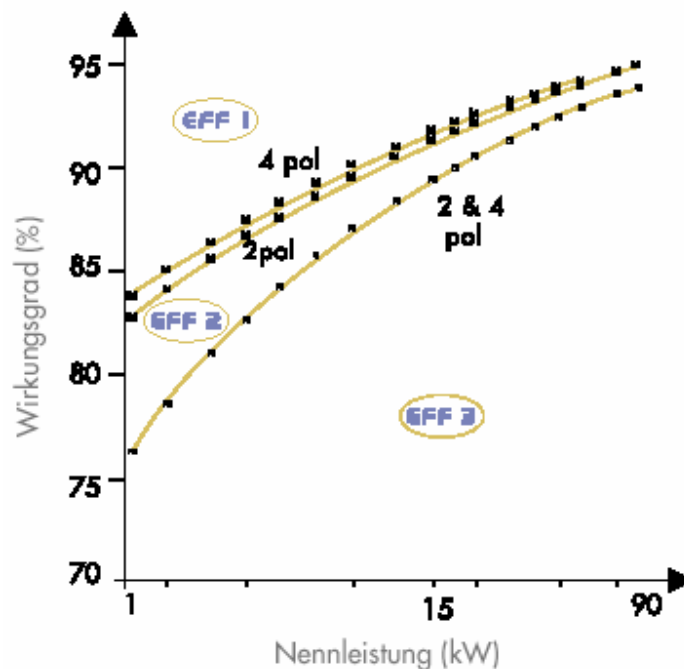


Abb. 2: Wirkungsgradklassen für Niederspannungs-Drehstrommotoren [3]

Die zweite Möglichkeit besteht darin, Frequenzumrichter (variable speed drives) vor den Motor zu schalten, um im Teillastbereich durch die Drehzahlregelung die Leistungsaufnahme zu reduzieren. Damit kann die mechanische Reduzierung, z.B. über ein Drosselventil, ersetzt werden. Die erreichbaren Energieeinsparungen erreichen in Abhängigkeit von den Lastfällen bis zu 50%. In der Abbildung 3 ist ein Berechnungsbeispiel für eine Pumpe bei einer max. geforderten Leistung von 17 kW dargestellt, wobei aus Sicherheitsgründen ein 22 kW Motor

gewählt wurde. Die obere Linie kennzeichnet die Leistungsaufnahme bei einer mechanischen Drosselregelung, die untere durch einen Frequenzumrichter.

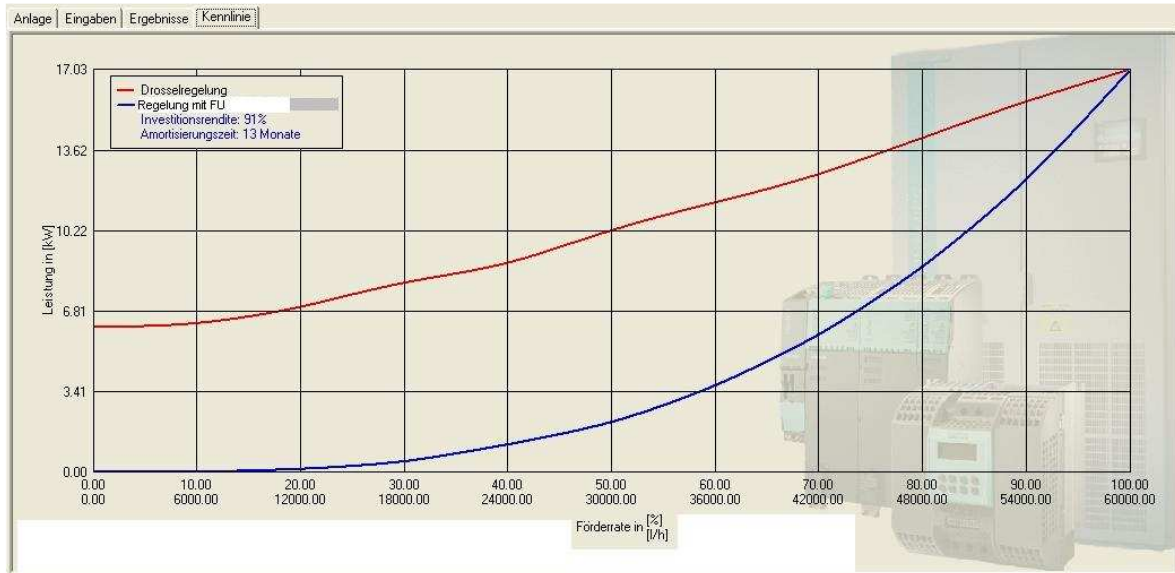


Abb. 3: Berechnungsbeispiel bei unterschiedlichen Durchflussmengen und mechanischer oder flexibler Drehzahlreglung durch Frequenzumrichter [5]

Als dritte Möglichkeit ist die Optimierung der Auslegung des gesamten angetriebenen Systems anzustreben, was nachfolgend allerdings nicht untersucht wird. Die Ausnutzung aller Möglichkeiten würde in der Nutzungsphase zu erheblichen Energieeinsparpotentialen entsprechend Tabelle 1 in Europa führen.

Tabelle 1: Überblick über das Energieeinsparpotential durch Motoren/ Motorensysteme in der EU [1]

	Einsparpotential (Billion kWh/ Jahr)					
	EU 15	EU 25	Frankreich	Deutschl.	Italien	Großbrit.
Energiesparmotoren	24	27	4	6	4	3
Frequenzumrichter	45	50	8	10	7	6
Optimierung der angetriebenen Systeme	112	125	19	26	17	15
Ges. Energieeinsparpotential	181	202	31	42	28	24

### 3 Ökologische Bewertung durch „Environmental Product Declaration- EPD“

Die Energieeinsparpotentiale durch höhere Wirkungsgrade oder durch Frequenzumrichter sind nur durch einen erhöhten Materialaufwand zu erreichen. Es wird hier untersucht, welche ökologischen Auswirkungen sich aus dem erhöhten Materialaufwand ergeben. Die ökologische Bewertung soll entsprechend der EuP-Richtlinie [2] nach der Methode der Ökobilanzierung nach ISO14041ff erfolgen. Bei dieser Methodik werden alle, während des Produktlebenszyklus verbrauchten Ressourcen sowie entstandenen Emissionen ermittelt, in Wirkungskategorien zusammengefasst und auf Leitsubstanzen (Äquivalente) normiert, um die Wirkungen in der Umwelt abschätzen zu können. Beispiele für Wirkungskategorien und deren Äquivalente sind:

- Global warming potential (GWP); Beitrag zum Treibhauseffekt in [kg CO<sub>2</sub> equ.]
- Ozon Depletion Potential (ODP); Ozonabbau in der Stratosphäre [CFCl<sub>3</sub> equ.]
- Acidification Potential (AP), Versauerungspotential [kg SO<sub>2</sub> equ.]
- Eutrophication Potential (EP), Eutrophierungspotential [kg Phosphate equ.]
- Photochemical Ozone Creation Potential (POCP), Bildung von Photooxidantien (Sommersmog) [kg Ethen equ.]
- Persistent Organic Pollutants (POP), langlebige organische Schadstoffe [ng i-Eq.]

Um die Methode der Ökobilanzierung zur Bewertung und zur Erstellung der EPD von Motoren zu nutzen, gibt es heute bereits eine Auswahl an angepassten Software-tools [z.B. 6,7,8]. Diese sind dadurch gekennzeichnet, dass

- nur relevante Lebenszyklusphasen für das spezielle Produkt betrachtet und
- nur nach Menge und Gefährlichkeit relevante Stoffe/ Stoffgruppen berücksichtigt werden.

Eigene Untersuchungen verschiedener Softwaretools zur Ökobilanzierung von Motoren haben dokumentiert, dass eine Normierung der Softwaretools unbedingt erforderlich ist, um dem Verbraucher wirklich vergleichbare Informationen zur Verfügung zu stellen. Bei gleicher Eingabe haben sich Schwankungsbreiten der Ergebnisse bis 94 % ergeben. Diese Vereinheitlichung muss nachfolgende Punkte umfassen:

- zu berücksichtigende Lebenszyklusphasen und Materialien, einschließlich zu treffender Verallgemeinerungen (z.B. Recyclingquoten)
- einheitliche Annahmen für Lastmodelle in der Nutzungsphase sowie Lebensdauer
- Auswahl der Kategorien der Wirkungsabschätzung und der Bezugsgrößen.

Nachfolgend soll anhand eines von der EU-geförderten Softwaretools [7] beispielhaft die ökologischen Auswirkungen auf den Treibhauseffekt durch den Einsatz

- von EFF1 Motoren und
- von Frequenzumrichtern

im Vergleich zu einem 22 kW EFF2-Standardmotor [6] dargestellt werden.

Alle Softwaretools basieren auf der Eingabe der Materialien (Bills of materials), die für das untersuchte Beispiel in der Tabelle 2 zusammengefasst werden. Für den EFF1-Motor wurde vereinfachend von einer 25% höheren Kupfermenge ausgegangen, die zu einer Erhöhung des Wirkungsgrades von 93% auf 95% führt.

Für einen, zum 22 kW-Motor passenden Frequenzumrichter (FU) wurde eine durchschnittliche Zusammensetzung gewählt.

Tabelle 2: Materialzusammensetzungen für EFF1 und EFF2 Motoren sowie Frequenzumrichter

		22 kW (EFF2)	22 kW (EFF1)	FU
Plastik	kg/kW	0,01	0,01	0,25
Eisenmetalle	kg/kW	11,36	11,36	3,57
NE-Metalle	kg/kW	0,73	0,88	1,32
Beschichtung/ Lack	kg/kW	0,14	0,14	0,04
Sonstiges/Elektronik	kg/kW	0,13	0,13	1,42
Summe in kg/kW		12,37	12,52	6,60

Für die Beurteilung der ökologischen Auswirkungen eines Frequenzumrichters sind die betrachteten Lastfälle des Motors von besonderem Interesse. Mit einer Polumschaltung ist nur eine stufenweise Reduzierung möglich. Bei dem ausgewählten stark schwankenden Lastfall kommt es zudem durch die variable Drehzahlregelung zu einer verringerten gleichmäßigeren Belastung des Motors und des Netzes. Für die Berechnungen wurde nachfolgender Lastfall (Tabelle 3) für den typischen Betrieb einer Pumpe gewählt. In diesem Fall kann durch den FU 40% des Strombedarfes in der Nutzungsphase gegenüber dem Netzbetrieb eingespart werden.

Tabelle 3: Lastfall für den Einsatz eines Frequenzumrichter bei Nutzung von 5000h/a

	Teillast	Erläuterung
3,3 h/d	30%	reduzierte Produktion
6,8 h/d	50%	Normalbetrieb
3,3 h/d	70%	erhöhte Produktion
0,3 h/d	100%	Spülen

Die Ergebnisse der Berechnungen mit dem VHK-Modell sind in der Tabelle 4 „Auswirkungen auf das Treibhauspotential“ bei einer Nutzungszeit von 15 Jahren und unterschiedlichen jährlichen Betriebszeiten zusammengefasst. Dort wird deutlich, dass sich auch bei geringen jährlichen Betriebszeiten (ab ca. 300h/a) die Nutzung von EFF1-Motoren und Frequenzumrichtern trotz des erhöhten Materialaufwandes hinsichtlich der Auswirkungen auf den Treibhauseffekt positiver darstellt (Ergebnis >0), als die Nutzung eines EFF2-Standardmotors. Diese Ergebnisse korrespondieren mit der erreichten Energieeinsparung, auch die Gesamtenergiebilanz ist bereits bei kurzen jährlichen Betriebszeiten positiv.

Tabelle 4: Vergleich der CO<sub>2</sub>-Äquivalente aus Herstellung und Nutzung eines 22kW-Motors über eine Nutzungsdauer von 15 Jahren bei unterschiedlichen jährlichen Betriebszeiten

	Standard 22kW	EFF 1 22 kW	Frequenz-umrichter	Standard - EFF1	Standard - FU
	Summe				
0 h/a *	36	45	61,2	-9	-25,2
300 h/a	2098	2066	1402	<b>32,24</b>	<b>696,5</b>
500 h/a	3473	3413	2123	59,74	1349,6
2000 h/a	13783	13517	8309	265,94	5473,6
5000 h/a	34402	33724	20681	678,32	13721,2

\* entspricht Aufwand der Herstellung

Die wirtschaftliche Amortisationszeit wird für die Annahme des Strompreises von 0,10 €/kWh in Abbildung 5 unten gezeigt. Bei 300 Jahresbetriebsstunden amortisiert sich der Einsatz des FU erst im 11. Jahr, so dass bei ökonomischer Betrachtung diese energetisch und ökologisch (bez. auf CO<sub>2</sub>-Emissionen, Abb.5 oben) sinnvolle Investition erst bei höheren Jahresbetriebsstunden durchgeführt würde (Amortisationszeit 19 Monate bei 2000 h/a).

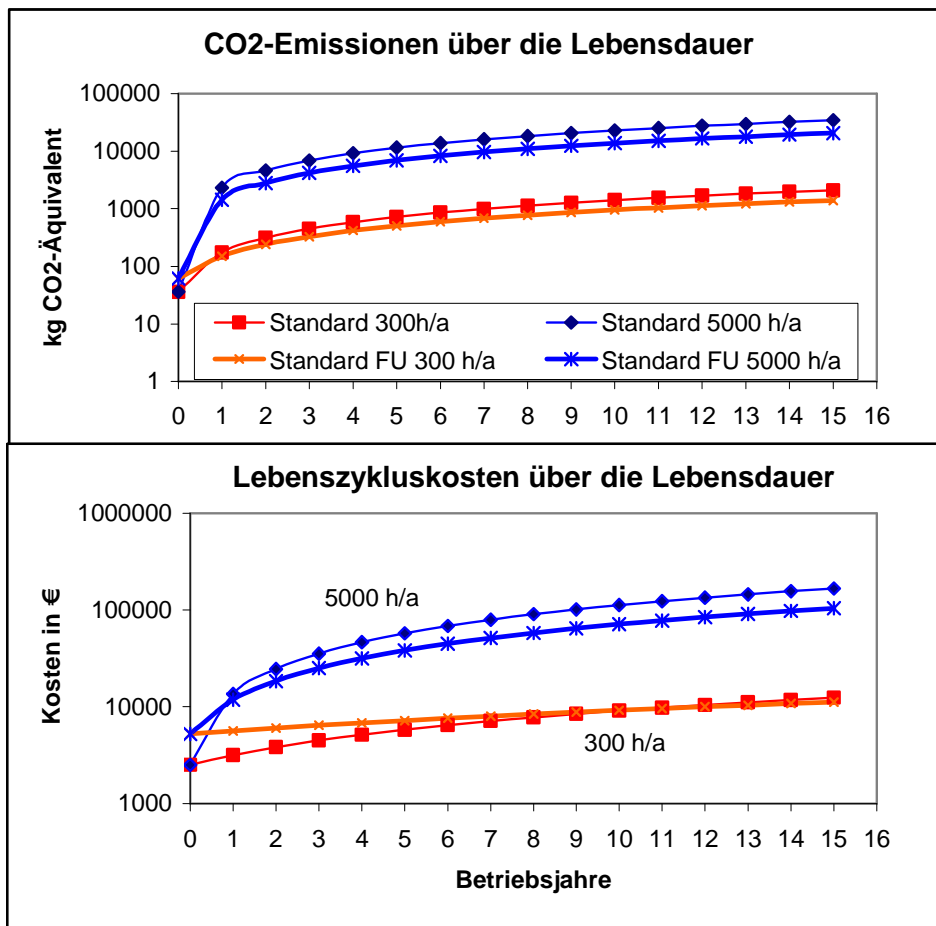


Abb. 4: Kohlendioxidemissionen und Gesamtkosten eines 22 kW Motors über die Lebensdauer von 15 Jahren



Obwohl die EuP-Richtlinie hauptsächlich die Energieeinsparung und die damit verbundene Reduktion der Treibhausgase adressiert, betrachtet die Methode der Ökobilanzierung weitere Wirkungskategorien, die zu abweichenden Aussagen führen können. Beispielweise werden in der Kategorie "Persistent Organic Pollutants" (POP) die zusätzlichen Aufwendungen resultierend aus der Herstellung für den FU bei 300 Jahresbetriebsstunden erst nach mehr als 15 Jahren ausgeglichen.

#### **4 Fazit**

Bedingt durch die sehr hohen Einsparpotentiale hinsichtlich energetischen und weiteren ökologischen Aspekten, die sich durch verbesserte Energieeffizienz bei Motoren ergeben können, werden auf europäischer Ebene verschiedene Anstrengungen unternommen, diese Einsparpotentiale zu erschließen. Neben verbesserter Gestaltung der Motoren oder der Entwicklung und Vermarktung von Frequenzumrichtern durch die Produzenten sollen auch die Nutzer verstärkt über die ökologisch vorteilhaften Antriebslösungen durch Ökoprotile/ Environmental Product Declaration informiert werden. Diese Information ist um so wichtiger, wenn die Einkäufer und die späteren Nutzer der Motoren nicht identisch sind [9]. Die zusätzlichen Ausgaben für den erhöhten Materialaufwand werden also nicht immer von den Nutzern, die die wirtschaftlichen und ökologischen Vorteile während der Betriebsphase erzielen, getragen. Es ist also erforderlich diese Nutzer zu sensibilisieren, so dass diese Aspekte z.B. in Ausschreibungen Berücksichtigung finden. Insbesondere weil die ökologischen Vorteile bereits früher tragen, als die ökonomischen. Um diese Information einfach und verständlich zu gestalten, sind Vereinheitlichungen sowohl bei den eingesetzten Softwaretools als auch bei der Darstellung der Ergebnisse auf europäischer Ebene erforderlich.

#### **5 Literatur**

- [1] De Keulenaer, H., Belmans, R.; Blaustein, E.; Chapman, D.; Almeida, A.; Radgen, P.: Energy Efficient Motor Driven Systems, in: EEMODS – Conference Proceedings, Vol.I, S. 170-180
- [2] Richtlinie 2005/32/EG des Europäischen Parlamentes und des Rates zur Schaffung eines Rahmens für die Festlegung von Anforderungen an die umwelt gerechte Gestaltung energiebetriebener Produkte und zur Änderung der Richtlinie 92/42/EWG des Rates sowie der Richtlinien 96/57/EG und 200/55/EG des Europäischen Parlamentes und des Rates
- [3] Sander, J.: Activities of CEMEP to Promote Efficient Motor Driven Systems, in: EEMODS 05– Conference Proceedings, Vol.I, S. 21-23
- [4] Heilmann, A.; Urban, U.: Ecological benefits of Energy-Saving Motors; in: EEMODS 05– Conference Proceedings, Vol.II, S. 208-213
- [5] Siemens AG – SinaSave 1.1, Stand 12.05 Programm zur internen Verwendung zur Berechnung der Energieeinsparungen durch drehzahlgeregelte Antriebe
- [6] [www.abb.com/motors&drives](http://www.abb.com/motors&drives)

A. Heilmann / U. Urban

[7] [www.vhk.nl/downloads.htm](http://www.vhk.nl/downloads.htm)

[8] [www.leonardo-energy.org](http://www.leonardo-energy.org);

[9] Sander, J.; Risch, W.: Energy Efficiency in Motor Driven Systems; in: Energy Efficient Motor Driven Systems, in: EEMODS 2005– Conference Proceedings, Vol.I, S. 170-180

D. Richter / H. Fritze / T. Schneider / P. Hauptmann / N. Bauersfeld / K.D. Kramer / K. Wiesner / M. Fleischer / G. Karle / A. Schubert

## Selektiver Gassensor für Hochtemperaturanwendungen auf der Basis piezoelektrischer Resonatoren

Denny Richter, Phone: ++49.3943-659-347, Fax: ++49.3943-659-5347, E-Mail: drichter@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

Holger Fritze, Hochschule Harz, Wernigerode

Thomas Schneider, Otto-von-Guericke-Universität Magdeburg,

Peter Hauptmann, Otto-von-Guericke-Universität Magdeburg,

Norman Bauersfeld, Hochschule Harz, Wernigerode

Klaus-Dietrich Kramer, Hochschule Harz, Wernigerode,

Kerstin Wiesner, Siemens AG, München,

Maximilian Fleischer, Siemens AG, München

Günther Karle, ProfilSchleif-, Fertigungs- & UmweltTechnik GmbH, Wernigerode

Andreas Schubert, ProfilSchleif-, Fertigungs- & UmweltTechnik GmbH, Wernigerode

### Kurzfassung

Resistive Gassensoren, welche allein die Änderungen der elektrischen Eigenschaften von Sensorschichten nutzen, zeigen bei hohen Temperaturen im Allgemeinen eine schlechte Gasselektivität. Das in dieser Arbeit vorgestellte Sensorprinzip auf der Basis hochtemperaturstabiler piezoelektrischer Langasitresonatoren ermöglicht die simultane Bestimmung der mechanischen und elektrischen Eigenschaften von Sensorschichten. Dadurch kann eine verbesserte Selektivität gegenüber verschiedenen Gasspezies erreicht werden. Zur Demonstration der Möglichkeiten des Sensorkonzeptes wurde die Unterscheidung von  $H_2$  und  $CO$  bei  $600\text{ °C}$  gewählt, wobei  $CeO_2$ -beschichtete Resonatoren eine deutliche Abhängigkeit des Messsignals von den vorliegenden Gasspezies aufweisen. Bei der simultanen Verwendung von  $TiO_2$ - und  $CeO_2$ -beschichteten Resonatoren lässt sich eine  $CO$ -Selektivität in  $H_2$ -haltiger Atmosphäre von 2 % abschätzen.

**Keywords:** Sensor, piezoelektrischer Resonator, Langasit

## 1 Einleitung

Auf Grund der wachsenden Bedeutung von Hochtemperaturprozessen in kleinen und mittleren Anlagen zur dezentralen Energieversorgung, wie Brennstoffzellenreformern und Kleinverbrennungsanlagen, steigt auch der Bedarf an Messsystemen zur Kontrolle und Regelung dieser Prozesse. Das erhöhte Interesse ist sowohl ökologisch als auch ökonomisch motiviert, da durch eine verbesserte Prozessregelung Schadstoffausstoß und Brennstoffverbrauch verringert sowie die Lebensdauer der Anlagen verbessert werden können.

Insbesondere im Bereich hoher Temperaturen fehlen jedoch kostengünstige Technologien und Verfahren zur selektiven Detektion von Gasen. So treten beispielsweise in der Hochtemperaturstufe von Brennstoffzellenreformern  $CO$ -Gehalte von 10 bis 15 % in einer  $H_2$ -haltigen Atmosphäre auf. Eine Unterscheidung von Kohlenmonoxid und Wasserstoff, wie sie zur Vermeidung von Katalysatorvergiftungen in PEM-Brennstoffzellen gefordert ist, könnte derzeit nur mittels optischer Technologien erfolgen, was auf Grund des hohen Preises bei kleineren und

mittleren Anlagen nicht vertretbar ist. Sensorsysteme, die auf den elektrischen Eigenschaften von Sensorschichten basieren sind einfach aufgebaut und daher kostengünstig. Sie bieten aber nur eine eingeschränkte Selektivität in Bezug auf reduzierende Gase wie  $H_2$  und  $CO$ .

Innerhalb eines BMBF-Projektes wird daher ein Sensorsystem entwickelt, bei dem neben den elektrischen auch die mechanischen Eigenschaften einer Sensorschicht mittels einer hochtemperaturstabilen piezoelektrischen Mikrowaage ermittelt und somit eine Erhöhung der Selektivität erreicht werden soll.

## 2 Sensoraufbau

Die Grundlage des Sensors bildet eine hochtemperaturstabile resonante Mikrowaage. Neuartige Piezoelektrika wie Langasit ( $La_3Ga_5SiO_{14}$ ) ermöglichen hierbei eine piezoelektrische Anregung bis zu Temperaturen von  $1470\text{ °C}$ , so dass sich hier neue Anwendungsmöglichkeiten im Vergleich zu herkömmlichen Materialien, wie z. B. Quarz, ergeben. Primäre Sensorelemente sind polierte und mit Platinelektroden versehene Langasitscheiben. Die Y-orientierten und ca.  $270\text{ }\mu\text{m}$  dicken Resonatoren zeigen eine Grundfrequenz von  $5\text{ MHz}$  bzw. eine Masseempfindlichkeit von  $50\text{ cm}^2\text{ Hz ng}^{-1}$ . Als Sensorschichtmaterialien kommen keramische Oxide in Frage, die eine Änderung ihrer Sauerstoffstöchiometrie in Abhängigkeit von der umgebenden Gasatmosphäre sowie eine gute chemische Stabilität aufweisen. Aussichtsreiche Materialien sind dabei  $CeO_2$  und  $TiO_2$ .  $1,5\text{ }\mu\text{m}$  dicke  $CeO_2$ - bzw.  $0,3\text{ }\mu\text{m}$  dicke  $TiO_2$ -Schichten wurden mittels Laser-ablation auf die Resonatoren aufgebracht.

Der prinzipielle Aufbau der verwendeten Resonatoren ist in Abbildung 1 im Querschnitt dargestellt. Die piezoelektrische Anregung erfolgt über zwei schlüssellochförmige Platinelektroden auf der Ober- und Unterseite des Resonators. Um den Einfluss unterschiedlicher Effekte auf das Resonanzverhalten bestimmen zu können wurden zwei verschiedene Elektrodenlayouts mit unterschiedlichen Elektrodendurchmessern auf Vorder- und Rückseite hergestellt.

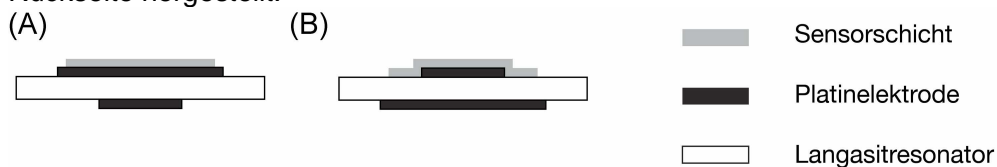
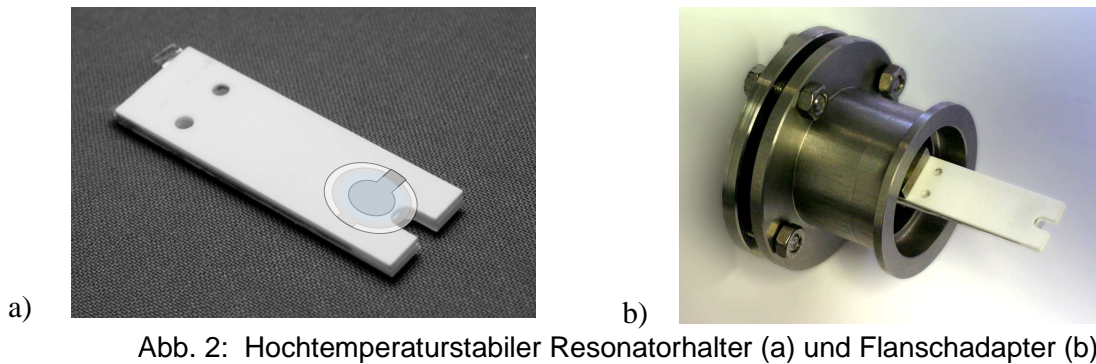


Abb. 1: Schematische Darstellung der verwendeten Resonatorlayouts im Querschnitt

Bei Layout (A) befindet sich die sensitive Schicht auf der Seite mit der größeren Elektrode. Es wirken sich hier nur Änderungen der mechanischen Eigenschaften bzw. der Masse der Sensorschicht auf den Resonator aus (Mikrowaagemodus). Bei Layout (B) hingegen befindet sich die sensitive Schicht auf der Seite mit der kleineren Elektrode und überlappt diese. Hier führen Änderungen der Schichtleitfähigkeit zu Änderungen der effektiven Elektrodenfläche, wodurch eine Verschiebung der Resonanzfrequenz verursacht wird (Leitfähigkeitsmodus). Eine detaillierte Beschreibung des Messeffektes findet sich in [1].

Mittels Ultraschallfräsen wurden hochtemperaturstabile Halterungen aus  $Al_2O_3$  hergestellt und zur Kontaktierung mit Platinsiebdruckschichten versehen. Zur Verbindung des Sensors an Teststände wird ein Standardflanschadapter verwendet (Abbildung 2). Gasentnahmesysteme entfallen auf Grund des Funktionsprinzips.



Die Resonatoren weisen bei hohen Temperaturen eine starke Dämpfung auf, so dass die Bestimmung der Resonanzfrequenz mit Hilfe einer Oszillatorschaltung kaum oder nicht möglich ist. Weiterhin enthält das vollständige Resonanzspektrum mehr Informationen als nur die Resonanzfrequenz. Aus diesen Gründen wurde ein spezialisierter Netzwerkanalysator entwickelt [2], der die Aufzeichnung der Impedanz  $\underline{Z}$  in der Umgebung der Resonanzfrequenz ermöglicht.

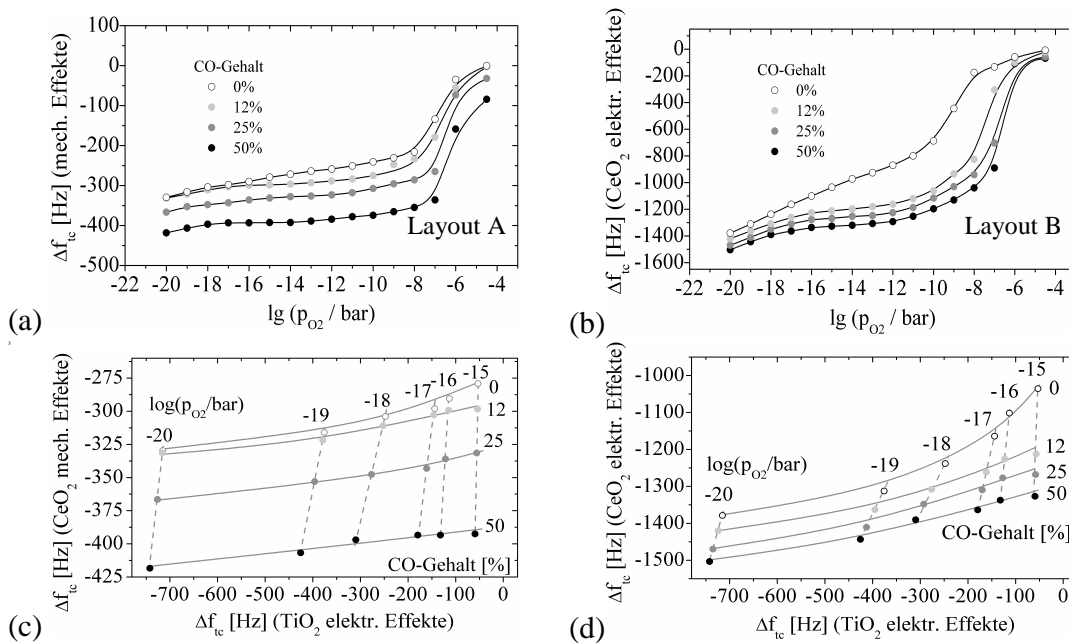


Abb. 3: Temperaturkompensierte Resonanzfrequenz CeO<sub>2</sub>-beschichteter Langasitresonatoren in verschiedenen Gasatmosphären mit bevorzugten mechanischen (a) bzw. elektrischen (b) Effekten sowie die Möglichkeiten der Gasselektion durch simultane Nutzung eines TiO<sub>2</sub>-beschichteten Resonators im Leitfähigkeitsmodus mit den beiden CeO<sub>2</sub>-Resonatoren.

### 3 Messungen

Die Bestimmung der Sensoreigenschaften erfolgte an einem Hochtemperaturmessplatz. Dieser besteht aus einem gasdichten Ofen, dem ein Gas definierter Zusammensetzung zugeführt werden kann. Die Messungen wurden in einem Sauerstoffpartialdruckbereich von  $10^{-5}$  bis  $10^{-25}$  bar bei einer Temperatur von 600°C durchgeführt. In einem Gesamtstrom von 20 cm<sup>3</sup>/min (Trägergas Argon) befand sich ein H<sub>2</sub>/CO-Gemisch mit einem Gesamtanteil von 0,25 %. Das Verhältnis von CO zu H<sub>2</sub> wurde dabei variiert (0 %, 12 %, 25 %, 50 % CO in H<sub>2</sub>).

Abb. 3a und 3b zeigen die temperaturkompensierte Resonanzfrequenz [3] der Resonatoren in Abhängigkeit des Sauerstoffpartialdruckes für die verschiedenen Gasgemische. Beide Resonatoren weisen eine deutliche Veränderung der Kurven bei unterschiedlichen Gasgemischen auf, wobei der Resonator mit dem Layout B vor allem bei niedrigem CO-Gehalt eine starke Differenzierung der Signale zeigt. Insbesondere kann bei dem Resonator, welcher vorwiegend die mechanischen Effekte der Sensorschicht widerspiegelt, im Bereich zwischen 12 % und 50 % CO eine ausgeprägte Sensitivität festgestellt werden.

Um die Möglichkeiten eines Sensors bei simultaner Nutzung von Resonatoren verschiedener Sensorschichten zu untersuchen, wurden die Messergebnisse der CeO<sub>2</sub>-beschichteten Resonatoren mit den Daten einer Messung mit TiO<sub>2</sub>-Sensorschicht kombiniert. Abbildungen 3c und 3d zeigen eine deutliche Trennung der verschiedenen Gaszusammensetzungen bei beiden Sensoreffekten. Anhand der Frequenzänderungen und des Rauschverhaltens des Sensors lässt sich gegenwärtig eine Genauigkeit der CO-Messung von etwa 2 % CO in H<sub>2</sub> auch für niedrige CO-Konzentrationen abschätzen, was den Anforderungen der Zielanwendung entspricht.

#### 4 Zusammenfassung

Der vorgestellte Sensor ermöglicht die simultane Bestimmung der elektrischen und mechanischen Eigenschaften einer Sensorschicht durch Nutzung unterschiedlicher Elektrodenlayouts, wodurch eine deutliche Verbesserung der Selektivität gegenüber reinen Leitfähigkeitssensoren erreicht wird. Bei einer Temperatur von 600 °C lässt sich anhand der bisher ermittelten Messdaten für CeO<sub>2</sub>- und TiO<sub>2</sub>-beschichtete Resonatoren eine Sensitivität des Sensors von 2 % CO in H<sub>2</sub> abschätzen.

Eine weitere Erhöhung der Empfindlichkeit ist bei Benutzung von Siebdrucksensorschichten zu erwarten, da diese zum einen dicker sind, zum anderen durch ihre Porosität eine größere effektive Oberfläche besitzen. Zudem sollen die Sensoren hinsichtlich ihrer Selektivität gegenüber anderen Gasen untersucht werden, um den Einsatzbereich des Systems auf weitere Anwendungen, wie Kleinfeuerungsanlagen, auszuweiten.

#### Literatur

- [1] Fritze, H.; Richter, D.; Tuller, H. L.: Simultaneous detection of atmosphere induced mass and conductivity variations using high temperature resonant sensors, *Sensors and Actuators B* 111-112 (2005) 200-206.
- [2] Schneider, T.; Richter, D.; Doerner, S.; Fritze, H.; Hauptmann, P.: Novel impedance interface for resonant high-temperature gas sensors, *Sensors and Actuators B* 111-112 (2005) 187-192.
- [3] Fritze, H.; Schneider, O.; Seh, H.; Tuller, H. L.; Borchardt, G.: High temperature bulk acoustic wave properties of langasite, *Phys. Chem. Chem. Phys.* 5 (2003) 5207-5214.

Die Autoren danken dem BMBF für die finanzielle Unterstützung des Projektes.

# Elektromechanische Eigenschaften von Langasitresonatoren bei hohen Temperaturen

Michal Schulz, Phone: ++49.3943-659-387, Fax: ++49.3943-659-5387, E Mail: mschulz@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

Holger Fritze, Phone: ++49.3943-659-346, Hochschule Harz, hfritze@hs-harz.de

## Kurzfassung

Langasit ( $\text{La}_3\text{Ga}_5\text{SiO}_{14}$ ) ist ein einkristallines, piezoelektrisches Material, welches als Grundlage für resonante Gassensoren oder Mikrowaagen bei hohen Temperaturen eingesetzt werden kann. Die Anwendungsgrenzen herkömmlicher Sensoren auf der Basis von Quarz werden durch die Verwendung von Langasit dank seiner sehr hohen Schmelztemperatur und seines guten Q-Faktors weit überschritten. Insbesondere weist Langasit keine Phasenumwandlung bis zur Schmelztemperatur von 1470 °C auf. Es konnte gezeigt werden, dass sich dieses Material noch bei 1400 °C zu Volumenschwingungen anregen lässt. Üblicherweise verwendete Y-Schnitt-Resonatoren weisen jedoch eine sehr starke Temperaturabhängigkeit der Resonanzfrequenz auf. Im Fall von Sensoren, die in einem größeren Temperaturbereich arbeiten müssen, kann eine Temperaturkompensation mit Hilfe des dritten Obertones erfolgen. Sollen die Sensoren jedoch bei einer annähernd konstanten Temperatur betrieben werden, ist ein temperaturkompensierter Schnitt von Vorteil. So kann auf eine nachträgliche Kompensation verzichtet werden. Derartige Sensoren lassen sich vorzugsweise in automatisierte Systeme integrieren. In diesem Artikel werden die elektromechanischen Eigenschaften von Langasit dargestellt. Die Impedanzspektren von Langasitresonatoren wurden bei Temperaturen bis 900 °C gemessen und zur Ermittlung der Materialdaten an die entsprechenden analytischen Lösungen der Bewegungsgleichungen gefittet. Anhand dieser Daten wurde ein Kristallschnitt vorgeschlagen, der eine minimale Temperaturabhängigkeit der Resonanzfrequenz für eine gegebene Temperatur aufweist. Bei 600 °C wird eine Verringerung der Temperaturabhängigkeit der Frequenz um den Faktor 500 erwartet.

**Schlagwörter:** Langasit, Resonator, Piezoelektrika

## 1 Einleitung

Langasit ( $\text{La}_3\text{Ga}_5\text{SiO}_{14}$ ) ist ein viel versprechendes Material für piezoelektrische Hochtemperaturanwendungen wie beispielsweise Gassensoren für Brennstoffzellen [1]. Eine Reaktion des Gases mit einer sensitiven Schicht führt zu einer Masseänderung, welche mit Hilfe von Langasitresonatoren bis 900 °C sehr genau bestimmt werden kann [2]. Neben der Temperaturstabilität bietet dieses Material sehr gute piezoelektrische und mechanische Eigenschaften wie beispielsweise einen hohen Gütefaktor und eine geringe Leitfähigkeit. Die Leitfähigkeitsänderungen beeinflussen die Resonanzfrequenz nicht, so lange der Sauerstoffpartialdruck höher als  $10^{-20}$  bar ist. Bemerkenswert ist, dass sich Langasit noch bei 1400 °C zu Volumenschwingung anregen lässt [3].

Die elektromechanischen Eigenschaften von Langasit wurden mit Hilfe der Impedanzspektroskopie untersucht. Anhand der Messungen bei niedrigen Frequenzen wurden der Volumenwiderstand und die dielektrischen Koeffizienten ermittelt. In diesem Fall wurde ein elektrisches Ersatzschaltbild für die Probe verwendet. Die Hochfrequenzmessungen in der Nähe der Resonanzfrequenz lieferten Daten, aus denen die piezoelektrischen und elastischen

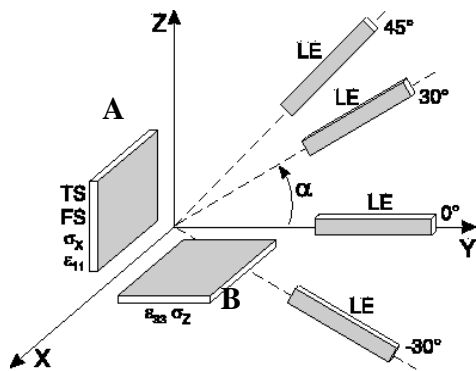


Abb. 1. Die Kristallschnitte und Kristallgeometrien, welche zur Bestimmung der elektromechanischen Eigenschaften des Langasits nötig sind [5].

Koeffizienten extrahiert werden konnten. Diese Messungen stellten ein erweitertes Resonanz-Antiresonanz-Verfahren dar, bei dem die aus den piezoelektrischen Gleichungen berechnete Impedanz eines Resonators mit den gemessenen Daten gefittet wurde. Die Berechnung geht von der Beschreibung der mechanischen Spannung  $T$  und der elektrischen Verschiebung  $D$  bei bekannter Dehnung  $S$  und elektrischem Feld  $E$  aus [5]

$$\begin{bmatrix} c & e^T \\ e & \epsilon \end{bmatrix} \cdot \begin{bmatrix} S \\ E \end{bmatrix} = \begin{bmatrix} T \\ D \end{bmatrix}. \quad (1)$$

Um die elastischen und piezoelektrischen Eigenschaften von Langasit vollständig beschreiben zu können, müssen die Komponenten  $c_{11}$ ,  $c_{12}$ ,  $c_{13}$ ,  $c_{14}$ ,  $c_{33}$  und  $c_{44}$  des Steifigkeitstensors und die Komponenten  $e_{11}$  und  $e_{14}$  der piezoelektrischen Koeffizienten ermittelt werden. Auf Grund der Kristallsymmetrie sind die verbleibenden Komponenten der genannten Tensoren gleich Null. Bei einigen Komponenten bietet es rechnerische Vorteile, die Kehrwerte  $s_{11}$ ,  $s_{12}$ ,  $s_{13}$ ,  $s_{14}$ ,  $s_{33}$  und  $s_{44}$  des Nachgiebigkeitstensors und  $d_{11}$  und  $d_{14}$  der piezoelektrischen Koeffizienten zu bestimmen. Der Einsatz von Resonatoren verschiedener Schwingungsmoden ermöglicht es, alle elastischen und piezoelektrischen Koeffizienten von Langasit zu ermitteln. An Hand der Kenntnis aller Koeffizienten des Materials bei verschiedenen Temperaturen kann ein Kristallschnitt derart ausgewählt werden, dass eine minimale Temperaturabhängigkeit der Resonanzfrequenz besteht.

## 2 Experimenteller Teil

Die Langasit-Einkristalle wurden mit Hilfe des Czochralski-Verfahrens im Institut für Kristallzüchtung, Berlin-Adlershof, gezüchtet [4]. Die benötigten Kristallgeometrien und Orientierungen [5] wurden laut Abb. 1 gewählt. Die vier Stäbchen (LE in Abb. 1) mit Abmessungen von 2 mm x 0,5 mm x 10 mm sind so orientiert, dass sie Längenschwingungen aufweisen und damit die Bestimmung der elastischen Nachgiebigkeitskomponenten  $s_{11}$ ,  $s_{13}$ ,  $s_{14}$ ,  $s_{33}$  und  $s_{44}$  ermöglichen. Eine Y-orientierte Platte (A in Abb. 1) wurde zu einer Flächenscherschwingung angeregt, welche durch die Nachgiebigkeit  $s_{44}$  beschrieben wird. Aus den Impedanzmessungen der Y- und Z-orientierten (B in Abb. 1) Platten wurden die Leitfähigkeit und die dielektrischen Koeffizienten ermittelt. Eine Dickenscherschwingung eines Y-Schnitt-Resonators wurde zur Bestimmung der Steifigkeit  $c_{66}$  verwendet. Anhand dieser Steifigkeit kann die Nachgiebigkeit  $s_{12}$  entsprechend  $s_{12} = s_{11} - (2c_{66})^{-1} - 2s_{14}^2/s_{44}$  berechnet werden. Im Fall der elastischen Koeffizienten wurden die Verluste in Form der Viskosität  $\eta$  berücksichtigt, welche als Imaginärteil der elastischen Komponenten dargestellt wurde:  $\mathbf{c} = c + j\omega\eta$  bzw.  $\mathbf{s}^{-1} = s^{-1} + j\omega\eta$  [5].

Die Proben wurden mit 200 nm dicken Platinelektroden versehen und im Temperaturbereich von 20 °C bis 900 °C untersucht. Die Leitfähigkeit und die dielektrischen Koeffizienten wurden mit Hilfe eines Impedanzspektrometers (Solartron SI 1260) und die elektromechanischen Eigenschaften mit Hilfe eines Netzwerkanalysators (HP 5100A) untersucht.



## 2.1 Impedanzmessungen

Die Impedanz der Y- und Z-orientierten Proben (A und B in Abb. 1) wurde mit einem elektrischen Ersatzschaltbild (Kondensator mit parallel geschaltetem Widerstand) gefittet, welches die Volumeneigenschaften der Langasitprobe bei höheren Temperaturen widerspiegelt. Der dielektrische Koeffizient, der Verluste im Dielektrikum in Form der Leitfähigkeit  $\sigma$  berücksichtigt, wird als komplexe Zahl  $\epsilon = \epsilon - j\sigma/\omega$  dargestellt.  $\omega$  ist hier die Kreisfrequenz. Der komplexe dielektrische Koeffizient wurde im Folgenden in einem physikalischen Modell des Resonators genutzt [5,7], um die elektrischen Verluste berücksichtigen zu können.

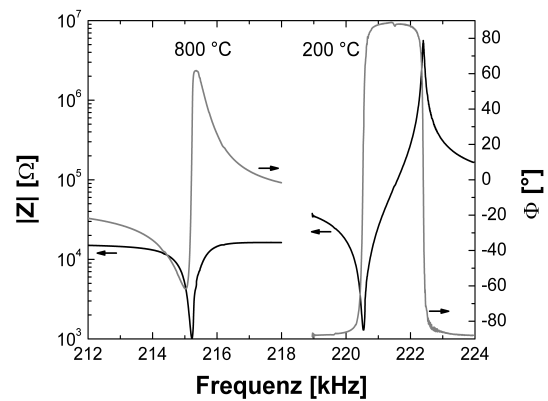


Abb. 2. Die Impedanzspektren eines Resonators bei 200 °C und bei 800 °C, welche in Längenschwingung angeregt wurde.

## 2.2 Längenschwingungen

Im Fall einer Längenschwingung (siehe Abb. 2) wurde die gemessene Impedanz unter Berücksichtigung der mechanischen und elektrischen Verluste an die folgende Impedanz, welche aus einem physikalischen Modell resultiert, gefittet [7]:

$$Z(\omega)^{-1} = Y(\omega) = j\omega \frac{\epsilon l w}{t} \left( 1 - \frac{d^2}{s \epsilon} + \frac{2d^2}{s \epsilon l w \sqrt{\rho s}} \tan\left(\frac{l w \sqrt{\rho s}}{2}\right) \right). \quad (2)$$

Hier sind  $l$  die Länge,  $w$  die Breite und  $t$  die Dicke des Resonators. Der piezoelektrische Koeffizient  $d$  und die Nachgiebigkeit  $s$  für eine beliebige Orientierung hängen vom Winkel  $\alpha$  des Stabes (siehe Abb. 1) ab und können aus der Tensorrotation berechnet werden:  $s = s_{11} \cos^4 \alpha + s_{33} \sin^4 \alpha - 2s_{14} \cos^3 \alpha \sin \alpha + (s_{44} + 2s_{13}) \cos^2 \alpha \sin^2 \alpha$  und  $d = -d_{11} \cos^2 \alpha + d_{14} \cos \alpha \sin \alpha$ .

## 2.3 Scherschwingungen

Im Fall einer Flächenscherschwingung ergibt sich die Resonanzfrequenz  $f_r$  eines Resonators mit der Kantenlänge  $l$  und der Dichte  $\rho$  wie folgt [5,6]:

$$f_r = \frac{\kappa}{\pi l} \sqrt{\frac{1}{\rho s_{44}}} \cdot \left( 1 - \frac{1}{2\kappa} \left( \frac{\kappa^2 - 2}{\kappa^2 + 2} \right)^{3/2} \sqrt{\frac{s_{11} + s_{33}}{2s_{44}}} \right), \quad (3)$$

mit  $\kappa \cong 2,0288$ . Die Nachgiebigkeit  $s_{44}$  wurde in diesem Fall aus der Resonanzfrequenz eines Resonators ermittelt.

Die Impedanz eines Resonators, welcher zu einer Dickenscherschwingung angeregt wurde, lautet wie folgt [7]:

$$Z(\omega) = \frac{t}{j\omega A_e \epsilon} \left( 1 - \frac{e_{26}^2}{c_{66} \epsilon} \cdot \frac{2}{\omega t \sqrt{\rho c_{66}^{-1}}} \tan\left(\frac{\omega t \sqrt{\rho c_{66}^{-1}}}{2}\right) \right), \quad (4)$$

wobei  $A_e$  die Fläche der Elektrode,  $e_{26}$  der piezoelektrische Koeffizient und  $t$  die Dicke des Resonators sind. Die hier resultierende Steifigkeit  $c_{66}$  berücksichtigt die Verluste.

### 3 Ergebnisse

Es wurden alle elastischen und piezoelektrischen Koeffizienten von Lagasit im Temperaturbereich von 20 °C bis 900 °C in Luft ermittelt. Die elastische Steifigkeit  $\mathbf{c}$  wurde aus der Inversion der Nachgiebigkeit bestimmt [7]:  $c_{ij} = (-1)^{i+j} \Delta_s^{ij} / \Delta_s$ , wobei  $\Delta_s$  die Determinante des Nachgiebigkeitstensors und  $\Delta_s^{ij}$  die Determinante des Nachgiebigkeitstensors ohne die Zeile  $i$  und Spalte  $j$  sind. Mit Hilfe der Tensorrotation wurde aus der Steifigkeit  $\mathbf{c}$  Resonanzfrequenz eines beliebig rotierten Resonators berechnet. Die Kenntnis der Temperaturabhängigkeit aller Parameter ermöglicht die Suche nach einem systematischen Parametervariation.

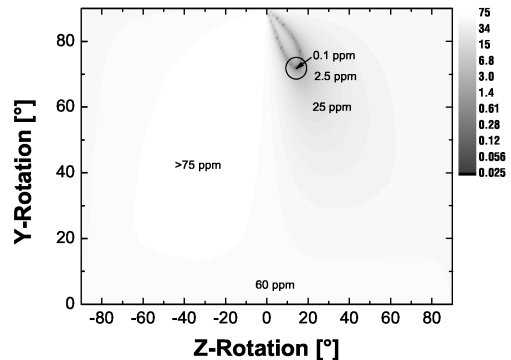


Abb. 3. Die relative Änderung der Resonanzfrequenz (in  $\text{ppm}\cdot\text{K}^{-1}$ ) eines Resonators als Funktion der Kristallorientierung.

temperaturkompensierten Schnitt durch

#### 3.1 Temperaturkompensierter Schnitt

Anhand des vollständigen Satzes der elektromechanischen Parameter wurde die Temperaturabhängigkeit der Resonanzfrequenz bei 600 °C als Funktion des Kristallschnittes für einen Dickenschwinger berechnet (siehe Abb. 3). Es wurde berechnet, dass im Fall eines Resonators, der 15° um die Z-Achse und 72° um die Y-Achse gedreht ist, der Temperaturkoeffizient der Resonanzfrequenz 500 mal kleiner als der Temperaturkoeffizient eines herkömmlichen Y-orientierten Resonators ist.

### 4 Zusammenfassung

Die elektromechanischen Eigenschaften von Langasit wurden im Temperaturbereich von 20 °C bis 900 °C mit Hilfe des Resonanz-Antiresonanz-Verfahrens bestimmt. Anhand der vollständigen Materialdaten konnte ein Kristallschnitt vorgeschlagen werden, der eine deutlich kleinere Temperaturabhängigkeit der Resonanzfrequenz im Vergleich zu herkömmlichen Resonatoren aufweist. Damit wird der Einsatz entsprechender Sensoren in automatisierten Systemen entscheidend vereinfacht.

### Literatur

- [1] Fukuda, T.; Takeda, P.; Shimamura, K.; Kawanaka, H.; Kumatoriya, M.; Murkami, S. et al., Proc. IEEE Int. Symp. on Applications of Ferroelectrics, (1998) 315-319.
- [2] Fritze, H.; Schneider, O.; Seh, H.; Tuller, H.L.; Borchardt, G., Phys. Chem. Chem. Phys., 5 (2003) 5207-5214.
- [3] Sauerwald, J.; Fritze, H.; Ansorge, E.; Schimpf, S.; Hirsch, S. Schmidt, B., International Workshop on Integrated Electroceramic Functional Structures, Berchtesgaden, Juni, 2005.
- [4] Ganschow, S.; Cavalloni, C.; Reiche, P.; Uecker, R., Proc. SPIE, 55 (1995) 2373.
- [5] Ikeda, T.: Fundamentals of Piezoelectricity, Oxford University Press, 1990, pp. 250-251.
- [6] Bechmann, R., Proc. Phys. Soc. (London), B64 (1951) 323-337.
- [7] Mason, W.P.: Physical Acoustics, Principles and Methods, Academic Press, 1964, p. 226.

Die Autoren danken der Deutschen Forschungsgemeinschaft für die finanzielle Absicherung dieses Projektes.

## Mikroelektromechanische Systemkomponenten für Hochtemperaturanwendungen

Jan Sauerwald, Phone: ++49.3943-659-389, Fax: ++49.3943-659-5-389, E-Mail: jsauerwald@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

Holger Fritze, Phone: ++49.3943-659-346, Hochschule Harz, hfritze@hs-harz.de

Erik Ansorge, Phone: ++49.391-67-12718, Otto-von-Guericke-Universität Magdeburg, erik.ansorge@e-technik.uni-magdeburg.de

Bertram Schmidt, Phone: ++49.391-67-18399, Otto-von-Guericke-Universität Magdeburg, bertram.schmidt@e-technik.uni-magdeburg.de

### Kurzfassung

Die Nachfrage in Bezug auf miniaturisierte Sensoren und Aktoren in Mess-, Steuer- und Regelprozessen steigt stetig mit neuen Einsatzgebieten. Unter harschen Bedingungen, wie beispielsweise hohen Temperaturen, fehlen oft noch geeignete Materialien oder Verfahren zur Realisierung von miniaturisierten Bauelementen. Auf der Basis von Langasit ( $\text{La}_3\text{Ga}_5\text{SiO}_{14}$ , LGS) werden ein Konzept und dessen Umsetzung in piezoelektrische mikroelektromechanische Systemkomponenten (MEMS) vorgestellt. Die materialwissenschaftlichen Grundlagen zur Herstellung miniaturisierter Strukturen werden dabei untersucht. Der Einfluss von oberflächennahen Dotierungen auf die elektrischen und chemischen Eigenschaften von LGS wird in dieser Arbeit mit dem Ziel untersucht, leitfähige Strukturen zu schaffen bzw. die Ätzgeschwindigkeit von nasschemischen Herstellungsprozessen zu modifizieren. Insbesondere werden die Diffusionskoeffizienten von Praseodym und Strontium in LGS bestimmt. Es kann eine Erhöhung der Ätzgeschwindigkeit nach thermischer Dotierung mit Praseodym gezeigt werden. Bei strontiumdotierten Proben wird eine Erhöhung der elektrischen Leitfähigkeit um den Faktor  $10^4$  beobachtet. Hergestellte Strukturelemente wie LGS-Biegebalken und Membranen können in einem Temperaturbereich von Raumtemperatur (RT) bis 750 bzw. 850 °C zu Schwingungen angeregt werden und zeigen akzeptable Gütefaktoren. Ein Entwurf für einen miniaturisierten Hochtemperaturgassensor auf Basis einer Hochtemperaturmembran wird vorgestellt.

**Schlagwörter:** Langasit, MEMS, monolithische Strukturen, Hochtemperaturanwendungen

## 1 Einleitung und Hochtemperaturkonzept

In vielen Bereichen wie Kraftwerken, Müllverbrennungsanlagen, der chemischen und der Halbleiterindustrie sind Prozesstemperaturen bis zu 2000 °C erforderlich. Aber auch im Bereich von Blockheizkraftwerken oder Brennstoffzellen findet man Temperaturen oberhalb von 600 °C. Dem großen Bedarf an hochtemperaturtauglichen Sensoren und Aktoren stehen nur wenige kostengünstige Lösungen gegenüber. Geeignete Materialien stellen die größte Herausforderung zur Übertragung von Sensorkonzepten in den Hochtemperaturbereich dar. Daher wird seit einigen Jahren an entsprechenden Sensor-Applikationen gearbeitet [1]. Für einen flächendeckenden Einsatz werden zudem kostengünstige und leicht handhabbare, also auch miniaturisierte, Systeme benötigt.

Auf der Grundlage eines neuartigen Konzeptes, d. h. der Schaffung von monolithischen Langasitstrukturen, werden mikroelektromechanische Systeme (MEMS) und Systemkomponenten für den Einsatz bei hohen Temperaturen entwickelt. Das

hochtemperaturbeständige piezoelektrische Material Langasit stellt das Substrat für Strukturierungen dar. Hervorzuheben ist, dass Langasit bis mindestens 1400 °C piezoelektrische Eigenschaften zeigt [2]. Zur Vermeidung thermischer Spannungen werden insbesondere monolithische Strukturen angestrebt. Eine daraus resultierende Zielstellung ist die Schaffung von leitfähigen Bereichen durch lokale Dotierungen. Diese Bereiche können beispielsweise zur Kontaktierung dienen. Des Weiteren muss bewertet werden, ob die Kontaktierungen langzeitstabil sind. Folglich ist die Kenntnis der atomaren Transportvorgänge bei hohen Temperaturen erforderlich. Durch gezielte Dotierung des Langasits mit Fremdatomen sollen die  $\text{La}^{3+}$ -Ionen durch zweiwertige bzw. vierwertige Ionen substituiert werden. Dadurch könnten Akzeptor- bzw. Donatorniveaus geschaffen werden, welche die elektrische Leitfähigkeit modifizieren. Zudem wird eine Änderung des chemischen Verhaltens beispielsweise gegenüber Ätzlösungen erwartet, so dass Möglichkeiten zur Strukturierung durch lokales Ätzen entstehen. Mikromechanische Systeme ausgehend von einfachen Strukturen wie Biegebalken oder Membranen über Mikropumpen bis hin zu integrierten aktiven elektronischen Elementen sind das Ziel dieser Entwicklungsarbeit.

Das Konzept wird mit Hilfe von drei Ansätzen verfolgt:

1. Bestimmung der Diffusionskoeffizienten der Dotieratome in Langasit zur gezielten Dotierung und Vorhersage der Langzeitstabilität von dotierten Bereichen.
2. Bestimmung der Leitfähigkeitsänderungen durch lokale Dotierungen.
3. Charakterisierung von piezoelektrischen MEMS im Hochtemperaturbereich.

## 2 Diffusionsuntersuchungen

Oberflächennahe Dotierungen wurden durch Aufbringen einer Quellschicht auf den LGS-Kristall und einem nachfolgenden Temperschritt erzeugt. Es zeigte sich, dass die Diffusion von Pr- bzw. Sr-Atomen in LGS relativ langsam ist. Die Diffusionskoeffizienten liegen in der gleichen Größenordnung. Mit  $D = D_0 \exp(-E/kT)$  ergibt sich für die Praseodymidiffusion:  $D_0 = 1,08 \times 10^{-4} \text{ m}^2/\text{s}$  und  $E = 4,5 \text{ eV}$ . Damit sind Dotiertiefen bis zu 10 µm bei Glühungen oberhalb von 1400 °C innerhalb praktikabler Zeiträume realisierbar. Bei 900 °C, d.h. bei der angestrebten Einsatztemperatur, sind die Diffusionskoeffizienten so klein, dass eine Langzeitstabilität erzeugter Strukturen gewährleistet ist. So vergrößert sich die charakteristische Tiefe eines Diffusionsprofils bei 900 °C von 3 auf 3,1 µm in  $4 \times 10^3$  Jahren.

Zur Erprobung von Strukturierungsprozessen wurden die Ätzgeschwindigkeiten von Langasit in Flusssäure und verschiedenen Ätzlösungen aus Salzsäure und Phosphorsäure in einem Temperaturbereich zwischen RT und 80 °C an dotierte n und undotierten Proben untersucht. Hier zeigte sich bei den Pr-dotierten Proben eine bis zu 40-fach gesteigerte Ätzgeschwindigkeit. Somit können beispielsweise mit lokalen Dotierungen Vertiefungen ohne Maskierungen während des Ätzprozesses erzeugt werden.

## 3 Leitfähigkeitsmessungen

Zur Ermittlung der Leitfähigkeit dotierter Bereiche wurden quadratische Proben mit einer Kantenlänge von 8 mm und einer Dicke im Bereich von 200 µm bis 1 mm auf einer Hälfte mit einer Quellschicht aus Praseodymoxid bzw. Strontiumoxid versehen und bei einer Temperatur von 1400 °C geglüht. Auf die Proben wurden mittels Photolithographie und Laserablation quadratische Elektroden sowohl im dotierten als auch im undotierten Bereich (Referenz) aufgebracht, welche Messungen der Leitfähigkeit im oberflächennahen Bereich zulassen. Mit

einem Impedanzspektrometer (Solatron SI 1260) wurden Spektren in einem Temperaturbereich von 200 bis 600 °C aufgenommen.

Die Temperaturabhängigkeit der Leitfähigkeit von Sr-dotierten LGS ist in Abbildung 1 dargestellt. Während bei den Pr-dotierten Proben keine Leitfähigkeitsänderungen in Bezug auf das undotierte Material beobachtbar war, zeigte sich bei den Sr-dotierten Proben eine signifikante Erhöhung der Leitfähigkeit im dotierten Bereich um 4 Größenordnungen. Eine Verwendung von dotierten Bereichen im Langasit als monolithische Elektroden wird geprüft.

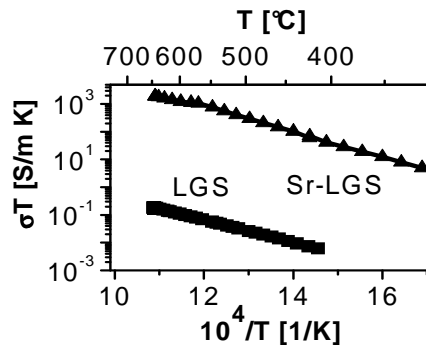


Abbildung 1: Temperaturabhängigkeit der Leitfähigkeit von Sr-dotiertem LGS

#### 4 Charakterisierung von Strukturelementen

Aus LGS-Substraten mit einer Dicke zwischen 300 und 800 µm wurden mit Methoden der Mikrostrukturierung wie nasschemisches Ätzen Biegebalken und Membranen herausgearbeitet. Die Kontaktierung erfolgte durch Titan-Platin-Elektroden auf der Ober- und der Unterseite. Die Membranen stellen im Wesentlichen verkleinerte eingespannte Resonatoren dar, welche wie konventionelle BAW-Resonatoren (Bulk Acoustic Wave) zu Scherschwingungen angeregt werden. Aufgrund der geringeren Dicke ergeben sich höhere Resonanzfrequenzen, welche mit einer angestrebten höheren Massensensitivität einhergehen. Gleichzeitig bleibt das System durch das umgebende Substrat stabil und handhabbar. Es wurden zwei verschiedene Herstellungsverfahren verwendet. Einerseits erfolgte die Strukturierung durch die Nutzung konventioneller Photolackmasken. Hier wurden Membranen bis zu einer Dicke von 80 µm gefertigt (Abb.2). Andererseits wurden Membranen mit Hilfe unterschiedlicher Ätzraten erzeugt. Dabei wurde die erhöhte Ätzgeschwindigkeit in Pr-dotiertem Langasit ausgenutzt, um beidseitig in einem runden gegenüberliegenden Bereich einen Ätzvorsprung von ca. 5 µm zu erzielen. Die so entstandenen Membranen wiesen eine Dicke von 230 µm auf. Die Charakterisierung der Membranen erfolgte mit einem Hochgeschwindigkeits-Netzwerkanalysator (HP E5100A). In Abbildung 3 ist die Temperaturabhängigkeit der Serienresonanzfrequenz und des Gütefaktors (Q-Faktor) einer Membran dargestellt.

In Analogie zu konventionellen BAW-Resonatoren zeigt sich eine Abnahme der Resonanzfrequenz und des Q-Faktors mit steigender Temperatur, welche auf eine Zunahme der



Abbildung 2: 80 µm dicke Membran

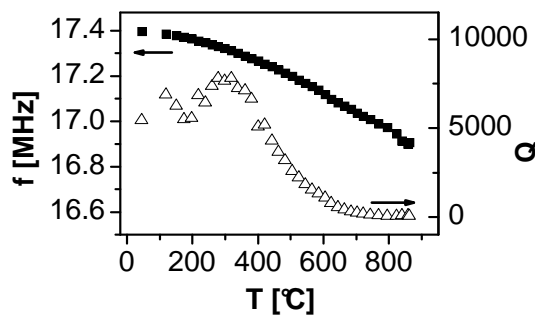


Abbildung 3: Abhängigkeit der Resonanzfrequenz und des Gütefaktors von der Temperatur

Dämpfung zurückzuführen ist. Mit dem Nachweis, dass Langasit-Membranen bis 850 °C zu Volumenschwingungen angeregt werden können, wurde gezeigt, dass miniaturisierte Mikrowaagen im Hochtemperaturbereich potentielle Bestandteile von MEMS sind.

## 5 Miniaturisierter Gassensor

Mit den in den Abschnitten 2 bis 4 vorgestellten Ergebnissen sind bereits die Voraussetzungen zur Herstellung eines miniaturisierten Gassensors gegeben. Die Grundlagen und die Funktionsweise des nachfolgend beschriebenen Gassensors sind in Arbeiten wie [3] detailliert erläutert. Analog zum scheibenförmigen Resonator einer konventionellen Mikrowaage werden aus einem LGS-Substrat eine oder mehrere Membranen herausgearbeitet. Dabei können die Membrandurchmesser aufgrund der reduzierten Membrandicke vergleichsweise klein gewählt werden. Eine Kontaktierung zur Anregung von Scherschwingungen erfolgte zunächst noch mit Platinelektroden. Auf einer der beiden Elektroden wird eine gassensitive Schicht abgeschieden. Durch die miniaturisierte Bauweise wird ein platzsparender Einbau in Sensorsystemen begünstigt. Weiterhin bieten Membranen grundsätzlich die Möglichkeit, mehrere Mikrowaagen auf einem Substrat zu betreiben. Diese können weitgehend ungestört schwingen, da sie voneinander entkoppelt sind. Arrays mit Sensorschichten bieten die Möglichkeit die Selektivität zu erhöhen.

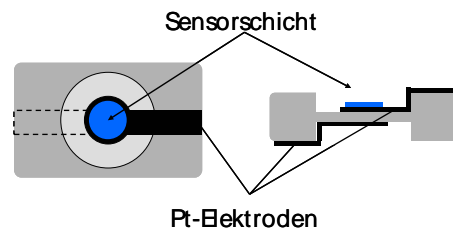


Abbildung 4: Schematischer Aufbau einer Sensormembran

## 6 Zusammenfassung

Die Herstellung von MEMS bzw. von entsprechenden Komponenten auf der Basis von Langasit wurde untersucht. Mittels oberflächennaher Dotierungen konnte eine Erhöhung der Leitfähigkeit erreicht werden, welche als Alternative zu metallischen Elektroden angesehen werden. In diesem Zusammenhang wurden die Diffusionskonstanten für Strontium und Praseodym in LGS bestimmt, so dass thermische Dotierungen kontrolliert werden können. Weiter zeigten dotierte Bereiche eine erhöhte Ätzgeschwindigkeit, so dass maskenloses Ätzen als eine weitere Möglichkeit zur Strukturierung zu Verfügung steht. Bei Strukturelementen wie Membranen und Biegebalken konnten Resonanzen und akzeptable Gütefaktoren bis zu Temperaturen von 800 bzw. 700 °C nachgewiesen werden. Die Übertragung eines Gassensorkonzeptes auf MEMS im Hochtemperaturbereich wurde vorgestellt.

## Literatur

- [1] Hornsteiner, J.; Born, E.; Fischerauer, G.; Riha, E.: Surface Acoustic Wave Sensors for High Temperature Applications. IEEE Int. Freq. Contr. Symp., (1998) 615-620.
- [2] Sauerwald, J.; Fritze, H.; Ansorge, E.; Schimpf, S.; Hirsch, S.; Schmidt, B.: International Workshop on Integrated Electroceramic Functional Structures, Berchtesgaden, 6.-8.6.2005.
- [3] Fritze, H.; Richter, D.; Tuller, H. L.: Simultaneous detection of atmosphere induced mass and conductivity variations using high temperature resonant sensors, Sensors and Actuators B 111-112 (2005) 200-206.

Die Autoren danken der Deutschen Forschungsgemeinschaft für die Unterstützung dieses Projektes.

# **Mobile und Geoinformationssysteme**





## Autonomous Mobile Robots for Outdoor Tasks

Prof. Dr. Harald Loose, Phone: ++49.3381-355-428, Fax: ++49.3381-355-499, E-Mail: loose@fh-brandenburg.de , Brandenburg University of Applied Sciences, Magedburger Straße 50, 14770 Brandenburg

Christiane Lemke, Phone: ++49.3381-355-427, Email: lemkec@fh-brandenburg.de , Brandenburg University of Applied Sciences

### Abstract

This paper presents a low-cost hardware platform for outdoor robots, being suitable for education, industrial prototyping and private use.

The choice of components is discussed, including platform, sensors and controller as well as GPS- and image processing hardware. Furthermore, a software approach is proposed, allowing students and researchers to easily implement own algorithms for localization, navigation and the tasks to fulfill.

Several robots can be integrated in a framework which connects various different hardware platforms, called the BOSPORUS network. Together with other components, they form an intelligent network for gathering sensor and image data, sensor data fusion, navigation and control of mobile platforms.

The architecture of a reference platform on the campus of the Brandenburg University of Applied Sciences is presented and evaluated.

**Keywords:** Navigation, Localization, Embedded Robotics, Image Processing, GPS, Mobile Systems, Communication Network

## 1 Introduction

The design, creation and programming of mobile robots is one of the key issues of mechatronics and makes up an important part in mechatronics education. A lot of robot navigation systems only work indoors. Outdoor operation comes with more challenges: fewer resources are available, the environment features like the lighting conditions are harder to control and modify, and operating conditions are more volatile[1]. But since outdoor robots are used in a lot of fields like demining or exploration and also offer the potential for new application areas, research and education dealing with them is very meaningful.

However, systems readily available on the market are usually very expensive. For educational, private and rapid prototyping purposes, cheaper platforms with more possibilities to develop and apply own algorithms are desirable. This paper presents an outdoor robot with the price of no more than 1500 Euro. It is based on a simple remote controlled vehicle and equipped with several low cost sensors: one GPS module with a gyroscope, a miniature color camera, two shaft encoders, seven ultrasonic range modules and seven infrared sensors. The main control over the vehicle is done by the RCUBE system, which is a platform for intelligent autonomous robots developed at Brandenburg University of Applied Sciences.

For successful fulfilling of tasks, localization and navigation are of vital importance. The robot needs to know where he is, where he wants to go and which path he will take. The use of a GPS module for navigation on the campus has been investigated and a GPS-map for this specific area was created. The results of the precision evaluations are presented. In the next step, various sensors, like shaft encoder, ultrasonic range modules, infrared distance sensors or robot vision and their role in localization are discussed along with the possibilities to fuse the acquired data and get the most possible precise information on the environment. The vision system has been evaluated in an object recognition task.

For researchers looking for a quick way to test their methods and for students, a well designed software approach is important. A Three-Layer-Architecture for the controlling unit is presented. The knowledge about the environment can be extended and completed if a network of mobile robots and sensors is created.

## 2 Hardware

To achieve the low price for the robot, the hardware components had to be chosen very carefully. The results are presented in this chapter and the components are described in terms of their function and precision.

### 2.1 Platform

An outdoor monster truck type Tamiya Trail Master QD 1/14 was chosen to be the first platform for the outdoor robot. It costs about 100 Euro and is available in toy shops.



Fig. 1: Monster truck platform, left: original, right: rebuilt

### 2.2 Controller

The RCUBE<sup>1</sup> system was developed in a project at Brandenburg University of Applied Sciences. It is a platform for intelligent autonomous mobile systems including vision, sensor/actuator and computing hard- and software. Its main features are a low power consumption, a number of ports for sensors and actuators exceeding those of systems on the market, a high computing power and on board image processing capability. It consists of 4 modules which can be connected by the 1Mbit CAN field bus.

A typical RCUBE system consists of four components: First, there is a CPU board with a 220 MHz StrongARM processor running ARMLinux, supplying the computing power of the system. Second, the VIO board makes up the standalone image processing module which is able to connect up to 4 standard PAL cameras. Third, the AKSEN board provides connections to robot components like actuators and sensors. Finally, a Compact Flash module gives a greater storage capacity.

The motion control of the autonomous mobile robot is based on the RCUBE platform [2][3]. Its components provide capability to sample sensor data, to control DC motors, to sample and process vision information, to communicate via an internal CAN bus and via a serial interface. It

---

<sup>1</sup> <http://ots.fh-brandenburg.de/rcube>

is also a key component of the BOSPORUS system and will be used, for example, for fixed and stand alone sensors, cameras and GPS receivers.

### 2.3 GPS

The Global Positioning System (GPS) will be used for localization of the robot. In theory, GPS positioning can achieve an accuracy of position measurements with deviations of less than one centimeter. In practice, standard hardware with reasonable prices has position deviations of several meters.

The robot is equipped with a non-expensive GPS receiver from the Swiss company u-blox<sup>2</sup> and costs less than 100 Euro. The characteristics of the board have been investigated with a series of measurements on a survey point on the roof of one of the university's buildings. In a period of 21 days, the GPS measurements of four receivers of four different points were sampled every 30 seconds. The statistics for two points measured over 24 hours are collected in table 1. It should be noticed, that the differences between the minimums and the maximums are about 33 m in the latitude (South-North-direction) and about 18 m in the longitude (West-East-direction), the deviations are about 4 m / 2 m. More details on the measurements can be found in [4]. Concluding, it can be stated that an autonomous mobile robot can be localized on the map using average data of a GPS receiver. The precision of the localization can be estimated by a few meters while separate measurements can differ more than 15 m. Averaged GPS data can be used directly for global localization and navigation, but the positioning error is too large for direct motion control. In the field of motion control, the incoming GPS data should be tested on the base of the probability that they are correct.

About 40 characteristic points of the campus were surveyed using the u-blox receivers, calculating the average of data sampled every second over two to five minutes. The coordinates of these points were determined according to a reference point and compared with the well-known coordinates of approximately the same points (fig. 2). The obtained deviations are less than 5 m for most of the points.

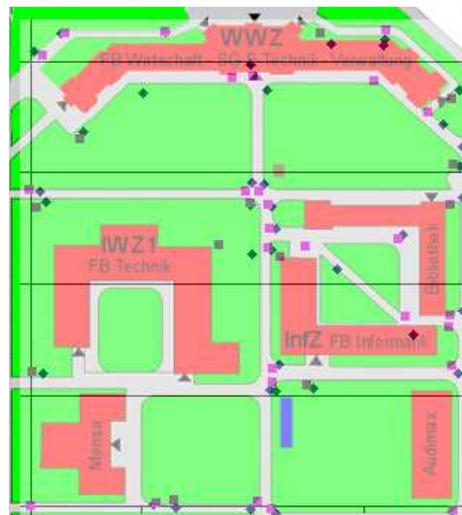


Fig. 2: Map of the campus  
(blue – measured, pink – well known points)

To improve the results of the GPS receivers, two main methods were pursued and tested. The first one is the usage of Differential GPS (DGPS), where the fact that two receivers close to each other are subject to the same error sources is utilized. The described methods have significantly improved accuracy achievable with GPS. The precision reaches up to 1.3 m in 2D and 2.3 m in 3D (with 95% probability). Detailed descriptions and analyses can be found in [5].

---

<sup>2</sup> <http://www.u-blox.de>

	Receiver 1		Receiver 2		Difference		Distance
	Latitude	Longitude	Latitude	Longitude	Latitude	Longitude	
Average:	5224.6281	1232.2275	5224.6274	1232.2268	1.37	0.60	2.33
Deviation	0.00216	0.00178	0.0017	0.00167			
[m]	4.00	2.01	3.15	1.89	1.77	1.08	1.05
Minimum	5224.621	1232.222	5224.623	1232.219	-3.86	-2.00	0.29
Maximum	5224.638	1232.238	5224.636	1232.232	5.28	3.17	5.28
Max-Min	0.0178	0.01571	0.01331	0.01325	9.14	5.18	4.99
[m]	32.98	17.77	24.66	14.99	9.14	5.18	4.99

Table I. Statistics of the data of two points and of their distance

The second implemented method of improving GPS measurements uses carrier phase information of the GPS signal. Usually, the position of a receiver is calculated with help of a code based measurement of so called pseudo-distances. More accurate algorithms observe the phase of the carrier signal, where an accuracy in the scope of milli meters is possible [6].

## 2.4 Sensors

The mobile robot positioning by the GPS receiver is supported by the fused data of various sensors (see figure 3). Different problems can be addressed with different types of sensors:

- Global positioning and orientation: GPS receivers, Gyroscopes
- Relative positioning with reference to known objects: image processing
- Relative positioning with regard to initial point: odometry
- Obstacle detection: ultrasonic, infrared or laser based sensors

One mobile robot is equipped 7 ultrasonic range modules, 7 infrared distance sensors and 2 shaft encoders, the other additionally with one GPS receiver connected to a gyroscope, one CCD camera.

GPS data helps to locate approximately the robot on a map of the campus. After knowing the area the robot is in, vision capabilities can be used to determine well known landmarks which are defined in advance.

The shaft encoders inside the wheels at the rear axis count rotations of the wheel and support dead-reckoning. Very small ultrasonic sensors measure distances to all objects that reflect the sent ultrasonic waves. Ideally, a range from 3 centimeters to 6 meters can be measured, but in practice, the analogue gain of the sonars had to be reduced for not catching unwanted signals from the floor. Right now, a distance up to 3.50 meters can be determined. Infrared sensors work after the same principle as sonars, but with infrared waves instead of ultrasonic. Sharp GP2Y0A02YK are used to measure distances in a range from 20 cm up to 150 cm.

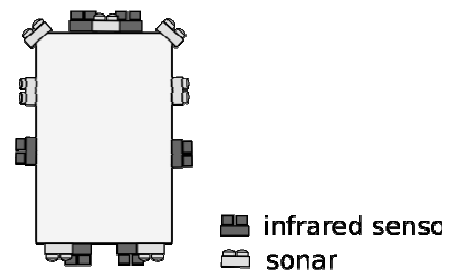


Fig. 3: Positions of Sonars and Infrared Sensors

## 2.5 Image Processing

The RCUBE platform provides sufficient performance to manage even the more complex tasks in the field of digital image processing in real time. An important role for navigation of the mobile robot plays image processing for landmark detection.

The implemented landmark detection system [7] works with the OpenCV library, which is optimized for real time computer vision and includes a collection of low-overhead, high performance operations to be performed on images. The challenge of image recognition for the embedded RCUBE system, which is of course not as powerful as an ordinary computer, was to find a good ratio between performance and robustness of the used algorithms.

First, the edges of a reference picture of the landmark are analysed. The best result from several investigated performant edge-detectors achieved a self-developed algorithm based on gray value differences. It outperformed the edge detection algorithm implemented in OpenCV as well as edge detection based on the SUSAN principle in processing time, number of correct detections and robustness.

Recognised edges are described using a gradient based descriptor in the next step. The edges are then analyzed and marked in the input picture, in which the object has to be found. Corresponding edges in both pictures are furthermore determined with a matching algorithm and then submitted to a Hough Transformation to eliminate false positives and to recognize the object. The time for a complete image recognition cycle is approximately three seconds.

## 3 Software Concept

Taking into account very different sources of information from the environment has the advantage, that the weaknesses of one source can be the strength of another, so that the reliability of the system is increased. But this comes with the price of a rising complexity of software design for such systems – building the hardware is just one part, using it productively is another important issue.

A robot navigating outdoors has to deal with an incomplete, inaccurate and approximate environment and the resulting “ubiquitous presence of uncertainty” which can not easily be modeled [8]. The idea of trying to model a complete environment has long been outdated by more modular approaches. One way are Three-Layered-Architectures based on three separate computational processes presented in [9]. The first proposed process is a reactive feedback control mechanism (“skill layer”), the second is a reactive plan execution mechanism (“sequencing layer”), and the last one is a mechanism for performing time-consuming computations (“planning layer”). Another way is to leave any kind of hierarchical approach and work exclusively with behaviors, a set of small independent units [10]. Each behavior is then alone responsible for fulfilling one specific task like following a line or detecting a landmark, some of which are suitable to be solved with fuzzy logic controllers.

In the Three-Layered-Architecture, higher level layers do not suppress the results of lower level layers, but provide input or advice. It is now applied to the different RCUBE modules.

### 3.1 Skill Layer

The skill layer, or the controller, is written in C language and is located on the AKSEN board. It is taking sensor input and computes direct commands for the actuators. Simple behaviors are implemented on this layer, like avoiding obstacles or following a wall. Knowledge about the world model is rarely necessary here.

### 3.2 Sequencing Layer

The sequencing layer knows a world model and thus the state of the mobile robot. It is written in C++ and runs on the CPU board of the RCUBE. It gives advice to the controller, in terms of which behavior is to use to what extent and to which time. Among others, it works with GPS input and a local representation of the campus. The computations of the sequencing layer can be more complex than those of the skill layer, but should be able to cope with bigger environmental changes.

### 3.3 Planning layer

The planning layer carries out time consuming computations. In this work, this layer is essentially used for image processing work for landmark detection and the process of smoothing carrier phase GPS data. It gives input to the sequencer by providing plans or it can answer queries from it. This layer is written in C++ and is located on the CPU board, taking its image data from the VIO board.

Different sources of information about the world and their interface to the RCUBE are shown in Figure 4.

### 3.4 Bosphorus integration

Single-robot systems can be used in a lot of scenarios. However, if several robots can be connected, they can fulfill more complex and challenging tasks, sharing their knowledge on the environment. Further components like fixed sensors, cameras and GPS receivers can additionally extend the workspace of a single robot. Controlling and visualizing devices like ordinary PCs or PDAs are also imaginable.

The goal of the developed system called BOSPORUS is to provide a generic approach to the omnipresent problem of exchange and processing of information.

The system uses peer to peer technology. It consists of three main components: one or more instances of the core (the super peer), other peers or clients and the platform-independent communication architecture.

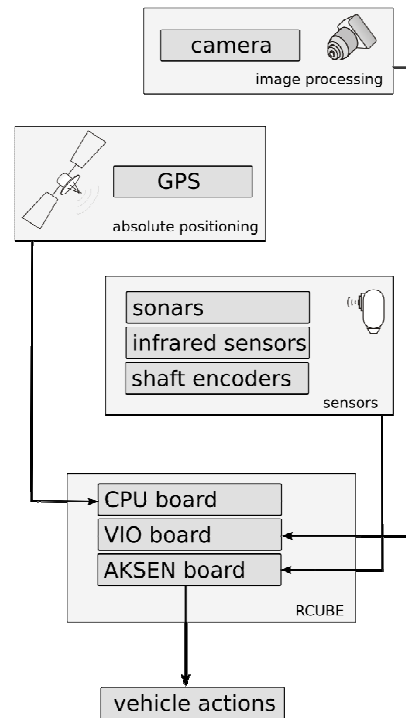


Fig. 4: Matching of environment input to RCUBE modules

A sample application demonstrating the capabilities of BOSPORUS has been developed (see Fig. 5 for illustration,). A PDA sends direct motion control commands to the robot, while motion is also simulated on a PC. Sensor values are sent from robot to PC and are visualized and evaluated there.

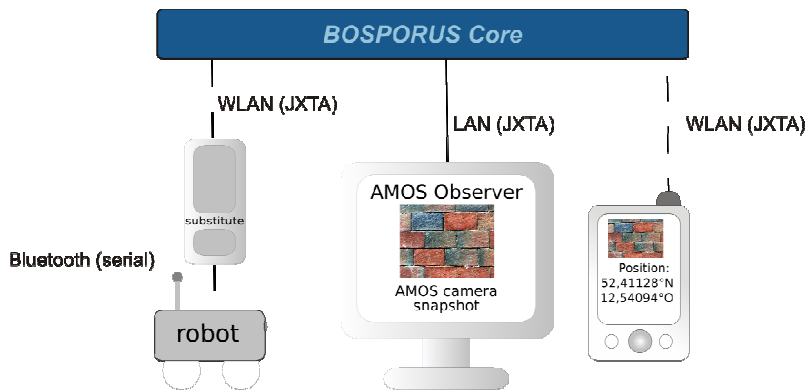


Fig. 5: Sample scenario of Bosphorus Network

## 4 Conclusion

This work describes a low-cost outdoor robot for educational and research prototyping use. The robot is controlled by the RCUBE platform with three main modules, providing actuator/sensor control, computing power and image processing capabilities.

The choice of non-expensive components is presented. Aside from sensors like sonars, shaft encoders and infrared sensors, a GPS board is used for navigation. Its characteristics have been evaluated and methods to improve the accuracy have been discussed. The navigation system is furthermore supported by landmarks recognized by an image processing system.

A three layer software concept has been applied the the RCUBE architecture, allowing parallel and highly reactive computations on the robot. The robot was also integrated in a Bosphorus network, allowing to create an intelligent network with other robots, cameras, PCs or PDAs.

## 5 Acknowledgement

The project BOSPORUS is supported by the "Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung" and the department of Computer Sciences and Media of the Brandenburg University of Applied Sciences.

## REFERENCES

- [1] S. Persa and P. Jonker, "Multisensor Robot Navigation Systems". In: Douglas W. Gage, Howie M. Choset (eds.), *Mobile Robots XVI (Proc. Conf. Boston, USA, Oct.28-Nov.2, 2001)*, Proc. SPIE, vol. 4573, 2002, 187-194
- [2] C. Lemke, H. Loose, I. Boersch, "Distributed Intelligent RCUBE Devices in Intelligent Space", Proceedings of the 6<sup>th</sup> International Workshop on Research and Education in Mechatronics, (2005), pp. 420-425
- [3] C. Lemke, H. Loose, I. Boersch: "Design of Mobile Platforms for a Distributed Intelligent Environment", Proceedings of the 9th International Conference on Mechatronics Technology in Kuala Lumpur, Malaysia, December 2005, pp. 54
- [4] H. Loose, C. Lemke, C. Papazov: "Image processing for navigation on a mobile embedded platform - Design of an autonomous mobile robot", In: CREUTZBURG, R. – TAKALA, J.: Proceedings of Multimedia on Mobile Devices II. SPIE Press, Conference Volume (to be printed in 2006).
- [5] D. Förder, "Verbesserte GPS-basierte Positionsbestimmung durch Verarbeitung der Rohdaten der Empfänger" (diploma thesis), Fachhochschule Brandenburg, August 2005
- [6] B. Hoffmann-Wellenhof, H. Lichtenegger, J. Collins, "GPS Theory and Practice", Springer Wien New York, 5<sup>th</sup> edition 2001
- [7] C. Papazov, "Objekterkennung auf der Plattform RCUBE" (diploma thesis), Fachhochschule Brandenburg, August 2005
- [8] D. Driankov, A. Saffiotti, "Fuzzy Logic techniques for autonomous vehicle navigation", Springer, Berlin, Germany, 2000.
- [9] E. Gat. On three-layer architectures. In D. Kortenkamp, R. P. Bonnasso, and R. Murphy, editors, *Artificial Intelligence and Mobile Robots*. MIT/AAAI Press, 1997. RR n3552 26  
Laugier, Fraichard, Garnier, Paromtchik and Scheuer  
<http://citeseer.ist.psu.edu/gat97threelayer.htm>
- [10] R. C. Arkin, "Behavior-Based Robotics", MIT press, Cambridge, MA 1998



## lejON – LeJOS Odometric Navigator

Florian Ruh, u20184@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

Jan Grohmann, u20189@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

Annedore Rößling, u20183@hs-harz.de, Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

### Abstract

Dieser Artikel skizziert das in [GRRS06] dokumentierte Projekt lejON. Darin wird die präzise Navigation von zweirädrigen Lego-Mindstorms-Robotern mit Hilfe der Odometrie<sup>1</sup> untersucht. Diese Roboter besitzen Rotationssensoren für ihre Antriebsräder und können in Java programmiert werden. Ziel der hier entwickelten Architektur namens lejON ist die Verbesserung und Erweiterung vorhandener Software zur odometrischen Navigation. Dazu stehen High-Level-Methoden zur Bewegungssteuerung bereit. Über lejON kann das Anwenderprogramm mit der Hardware kommunizieren. Die Steuerung übernimmt dabei ein Zustandsautomat. Ein softwareseitiger PID-Regler<sup>2</sup> wirkt den in der Praxis auftretenden Störfaktoren entgegen. LejON unterstützt verschiedene Fahrzeugarchitekturen, u. a. das (*Single*) *Differential Drive* (DD). Testläufe mit einem DD-Prototypen zeigen, dass die Odometrie der Roboter hinreichend genau mit der tatsächlich gemessenen Position übereinstimmt. Die lejON-API ist also nicht nur eine Verbesserung der bisherigen Navigation sondern bietet darüber hinaus eine höhere Genauigkeit sowie neue Funktionen.

**Keywords:** Navigation, Odometrie

## 1 Konzept

Dieser Abschnitt stellt das Konzept einer modularen Software-Architektur zur odometrischen Navigation zweirädriger *wheeled mobile robots* (WMRs) auf Lego-Mindstorms-Basis mit LeJOS-Kernel vor. Es trägt den Namen lejON – LeJOS Odometric Navigator. LeJOS selbst ist ein Open-Source-Projekt, welches die Lego-Firmware auf dem Mindstorms-Baustein mit einer Java Virtual Machine ersetzt. Die Ziele bei der Entwicklung dieser Architektur sind:

1. Abstraktion eines zweirädrigen WMRs auf DD-Basis in eine Navigatorschicht – der lejON Schicht – mit einer erweiterbaren API.
2. Implementierung einer rein odometriebasierten Positionsschätzung.
3. Bereitstellung von Funktionen zur Realisierung von parametrisierbaren Figuren: Translation, Rotation und Kurvenfahrt.
4. Implementierung eines softwarebasierten Regelmechanismus zur Kontrolle und Korrektur der Radbewegungen.

### 1.1 Odometrie

Odometrie (griech. *hodos*: der Weg, *metron*: das Maß) ist die Messung von Wegen und Winkel anhand der Radbewegungen. Hierfür werden Rotationssensoren verwendet, welche die

---

<sup>1</sup> Odometrie: Weg- und Winkelmessung

<sup>2</sup> PID: Universeller Regelungsbaustein (vgl. Automatisierungstechnik)

Umdrehungen der Räder um ihre Achspunkte messen. Die Sensoren arbeiten als inkrementelle Kodierer, welche den Drehwinkel der Sensorwelle auf einen Code abbilden, vgl. [CH05]. Deshalb spricht man auch vom *shaft encoder*.

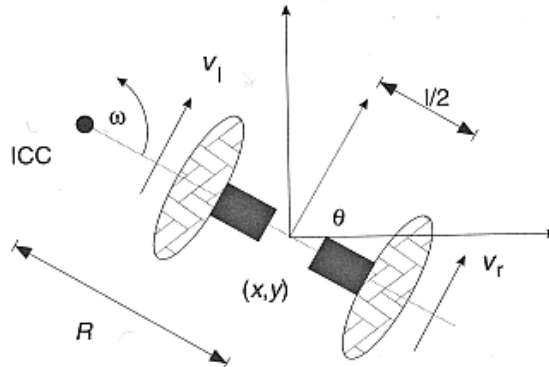


Abb. 1: DD-Modell im Koordinatensystem  
Quelle: [DJ00]

Die Position eines sich in Bewegung befindenden WMRs ist über seine Vorwärtskinematik bestimmt und kann im Idealfall durch Odometrie exakt berechnet werden. Praktisch treten jedoch systematische Messfehler und externe Störungen auf, welche die exakte Navigation erschweren.

## 1.2 lejON-Schicht

In [KBM98, S.93,103ff.] wird eine abstrakte Schichtenarchitektur für mobile Indoor-Roboter beschrieben, welche sich grob in *High-Level-Planner* und *Low-Level-Navigation&Control* unterteilt. Dieses Konzept nutzt einen *Command/Report-Mechanismus* als Schnittstelle zwischen den Schichten, welche nach unten hin von den Details der Aktorik und Sensorik abstrahiert, nach oben hin Sensor- und Statusinformationen in aufbereiteter Form bereitstellt. Die Adaption für dieses Projekt wird aufgegriffen und führt zu dem in Abb. 2 dargestellten Architekturmodell.

**Schnittstellen** Die Kommunikation zwischen lejON und der Anwendung erfolgt über Nachrichtenblöcke. Dieses sog. *Motion Command* definiert eine der Fahrfiguren "Translation", "Rotation", "Kurve". Eine Folge solcher Kommandos beschreibt dann einen aus Segmenten bestehenden Pfad. Im *Motion Report* wird die geschätzte eigene Position zusammen mit weiteren Statusinformationen bereitgestellt. Diese ist relativ zur Startposition der aktuellen Figur.

**Motion Controller** Die internen Aktionen vor, während und nach der Ausführung eines *Motion Commands* werden durch einen Zustandsautomaten nach Abb. 3 koordiniert, der hier als *Motion Controller* bezeichnet wird.

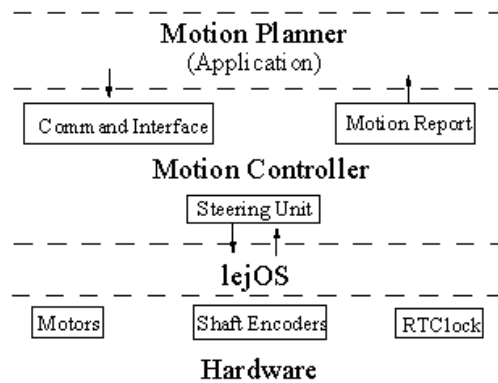


Abb. 2: Architektur der leJON-Schicht

**Steering Unit** Die Schnittstelle zum LeJOS-Kernel und der Hardware des WMRs ist die Steering Unit. Sie stellt ein internes API zur Verfügung, welches in den Zuständen des Motion Controllers verwendet wird. Hier werden alle Aufgaben und Funktionen gekapselt, die mit der hardwareseitigen Vorbereitung, Durchführung und Terminierung einer Figur verbunden sind. Insbesondere die Positionsschätzung und die Fehlerkompensation sind hier implementiert. Ohne die eigentliche Realisierung vorweg zu nehmen, ist es notwendig, die beiden letztgenannten Aufgaben als periodische Tasks mit hinreichend hoher Frequenz zu realisieren.

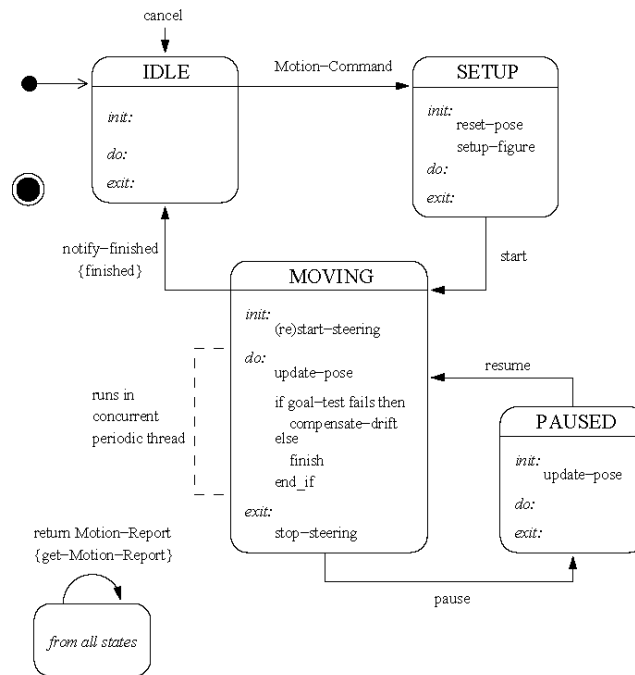


Abb.3: Zustandsdiagramm des Motion Controller

Die leJON-Architektur ist unabhängig vom WMR-Typ. Der typspezifische Teil dieser Architektur ist lediglich die Steering Unit. Im Folgenden wird vom DD als WMR-Typ ausgegangen.

## **2 Architektur**

Die Architektur des zweirädrigen Roboters für dieses Projekt soll eine sichere und zuverlässige Navigation erlauben. Dafür ist es erforderlich, möglichst viele Störfaktoren bereits in der Planungsphase zu beseitigen. Insbesondere soll der Roboter ohne Drift geradeaus fahren können. Dies ist bei herkömmlichem Antrieb über zwei Motoren, die mit jeweils einem Antriebsrad verbunden sind, ohne Software-Regelung nicht möglich. Hierbei spielen die unterschiedlichen Motorenbeschaffenheiten eine entscheidende Rolle.

Die für dieses Projekt realisierte Architektur ist das (Single) Differential Drive.

### **2.1 Architektur eines Differential Drives**

Bei einem DD werden die Antriebsräder nicht miteinander verbunden sondern von jeweils einem Motor betrieben. Infolgedessen ist sowohl die Geradeaus- als auch die Kurvenfahrt von beiden Motoren gleichermaßen abhängig. Dieses Verhalten ist problematisch, da durch kleinste Abweichungen im Fertigungsprozess, durch Materialfehler oder durch Verschleiß nicht gewährleistet werden kann, dass sich – trotz gleichmäßiger Ansteuerung – die Motoren mit exakt demselben Drehmoment drehen.

### **2.2 DIDI als Prototyp eines DD**

Der in diesem Projekt zum Einsatz kommende Prototyp, benannt DIDI (Abk. für Differential Drive), basiert auf einer DD Architektur.

Die getrennt voneinander arbeitenden Antriebsräder werden von jeweils einem Motor über Zahnräder angetrieben. Die Rotationssensoren sind mit den Antriebsachsen ebenfalls über Zahnräder verbunden. Die Übersetzungsverhältnisse Motor zu Antriebsachse und Antriebsachse zu Rotationssensor können auf einfache Weise verändert werden. Das Stützrad befindet sich in einiger Entfernung zum Fahrzeugschwerpunkt. Seine Aufhängung kann sowohl in der Höhe als auch in der Entfernung zum Schwerpunkt leicht geändert werden. Dies ist dann besonders sinnvoll, wenn z. B. die Antriebsräder durch kleinere ersetzt werden.

### **2.3 Einschränkungen und Berücksichtigungen**

Die o. g. allgemeine Architektur kann nicht ohne Weiteres mit Lego-Technic-Bauteilen konstruiert werden. Hierzu müssen die Grenzen der eingesetzten Materialien berücksichtigt werden. Da sich ein Kettenantrieb mit Lego aus anderen Projekten heraus als nicht geeignet für eine präzise Navigation herausstellt, ist für dieses Projekt ein dreirädriger Roboter vorgesehen. Somit muss das Design aus [Bau00] an drei Räder angepasst werden. Dabei ist zu beachten, dass der Roboter über zwei Räder angetrieben wird und das dritte als Stützrad nutzt. Letzteres muss aber von zu großer Last befreit werden, damit seine Drehung bei Richtungsänderung des Roboters nicht die gesamte Konstruktion dreht und somit die Fahrtrichtung negativ beeinflusst. Aus diesem Grund ist der schwerste Baustein, der RCX, direkt über der Antriebsachse angebracht, so dass er auf das Stützrad möglichst keinen Druck ausübt.

Das Stützrad lässt sich somit fast frei bewegen und passt sich (passiv) der Bewegungsrichtung des Roboters an.

Das Eigengewicht ist allein durch den RCX und die Batterien sehr hoch. Deswegen sollte darauf verzichtet werden, den Roboter mit übermäßig breiten Reifen oder mit hoher Profilstärke fahren zu lassen, um möglichst wenig Reibung auf dem Untergrund zu erzeugen.

An jedem Antriebsrad ist ein Rotationssensor angebracht. Da allein schon die Übersetzung durch Zahnräder mit einem gewissen "Spiel" verbunden ist, wird das präzise Auslesen bei geringen Geschwindigkeiten besonders wichtig.

Viele Störfaktoren lassen sich somit schon während der Fahrzeugkonstruktion minimieren. Die exakte Geradeausfahrt ohne Kursabweichung, der Idealfall, lässt sich dennoch nicht realisieren. Ausschlag gebend sind die Kunststoffbauteile, wobei sich gerade die Radachsen stark durchbiegen und das Fahrzeug dadurch ins Schlingern bringen. Somit ist eine Navigation nicht ohne Software-Regelung möglich.

### 3 Evaluation

Zur qualitativen Beurteilung der Navigation mittels lejON-API werden die bereitgestellten Fahrfiguren getestet und die Ergebnisse mit äquivalenten Figuren unter Einsatz des *RotationNavigators*, eines LeJOS-Bestandteils, verglichen. Als Architektur dient hierbei der DIDI-Prototyp. Im Wesentlichen soll geklärt werden, inwieweit eine vorgegebene Figur vom Fahrzeug gefahren wird, wie groß die gemessene Abweichung vom Zielpunkt ist und wo das Fahrzeug glaubt sich zu befinden.

Hierbei wurde festgestellt, dass insbesondere die Geradeausfahrt gegenüber der herkömmlichen Variante des *RotationNavigators* verbessert werden konnte. Auch im Bereich der Rotation gibt es gewisse Vorteile, obwohl in den Tests der *RotationNavigator* hier mit geringeren Fehleraufschlägen rotiert. Jedoch ist es mit dieser LeJOS-Navigatorklasse nicht möglich die Position während der Fahrt genau zu bestimmen. Bei einer Geradeausfahrt werden beispielsweise Drift und Richtungsänderung überhaupt nicht registriert. Hier liegt der große Vorteil der lejON-API. Es kann nicht nur die Änderung der x- und y-Koordinate relativ genau vom Fahrzeug ermittelt werden sondern auch zusätzlich die Richtungsänderung. Mit diesen Daten kann der Anwender den Restfehler einer Fahrfigur ausgleichen oder berücksichtigen. Die Fehlertoleranz für die seitliche Kursabweichung liegt für lejON und den hierfür benutzten Prototypen bei einem 9° breiten Korridor. Aufgrund des Schwingens durch die Regelung ändert sich auch die Ausrichtung des Fahrzeugs ständig. In den bisherigen Testläufen lag die Richtungsänderung zum Bezugskordinatensystem im Durchschnitt bei  $\pm 8^\circ$ , im Maximum bei  $\pm 13^\circ$ . Bei den Messreihen zur Rotation (vgl. Tabelle 1) offenbart sich ein Offsetfehler, der zwischen 25° bis 50° liegt. Die Ursachen hierfür sind wiederum in Architektur und der Auflösung der Rotationssensoren sowie der geringen Anzahl an vorhandenen Motorleistungsstufen zu suchen. Weiterhin ist zu bedenken, dass beim Motorenstopp das Fahrzeug aufgrund seiner Trägheit weiterrollt. Bei einem Nachrollen von 3 TICKS<sup>3</sup> und einer Genauigkeit von 9 TPD<sup>4</sup> pro Sensor entspricht dies einem Fehler von 54°. Positiv ist aber, dass die Software den Trägheitseffekt registriert und in der Positionsberechnung diesen Fehler berücksichtigt. Im Gegensatz hierzu sind die Werte (vgl. Tabelle 2) beim *RotationNavigator* beim Drehen genauer, jedoch kann die Software den tatsächlich gefahrenen Winkel nicht feststellen.

---

<sup>3</sup> TICKS: Impulse des Rotationssensors

<sup>4</sup> TPD: TICKS pro Grad in Bezug auf die Ausrichtung des Fahrzeugs

Tabelle 1 Mittelwerte der Drehwinkel. Links befinden sich die gemessenen Werte (M), rechts die vom Fahrzeug berechneten (B). Fünf Messungen pro Leistungsstufe.

	Rotation (lejON)							
	45°		90°		180°		360°	
Leistungsstufe	M	B	M	B	M	B	M	B
0	68,4	71,6	133,4	133	209,4	216,4	386,4	395,6
1	68,4	71,8	135,6	134,4	201,2	203	385,6	394,2
2	84,4	84,2	120	121,2	212,2	213	395,6	405,4
3	95,4	96,8	132,4	136,6	222,6	227,6	382,6	388,6
4	94	98	143	147,2	226,6	225,6	388	397,4
5	74,2	77,2	114	116	210,2	213,2	395,4	406,6
6	69,8	68,2	110	114	208	212,2	406,4	417,4
7	74,4	74,4	115,4	116	208,6	214,8	397,8	409,2

Tabelle 2 Mittelwerte der Drehwinkel mittels *RotationNavigator*. Links befinden sich die gemessenen Werte (M), rechts die vom Fahrzeug berechneten (B). Drei Messungen pro Leistungsstufe (n.e. – nicht evaluiert).

	Rotation (RotationNavigator)							
	45°		90°		180°		360°	
Leistungsstufe	M	B	M	B	M	B	M	B
0	148,33	45	n.e.	n.e.	n.e.	n.e.	n.e.	n.e.
1	50	45	95	90	213,3	180	385	360
2	50	45	98,3	90	188,3	180	341,6	360
3	53,3	45	101,6	90	180	180	331,6	360
4	55,67	45	100	90	183,3	180	330	360
5	55,67	45	101,67	90	181,7	180	330	360
6	59	45	103,33	90	181,7	180	331,7	360
7	58	45	99	90	180	180	338,3	360

Ein weiterer Vorteil der lejON-API ist die Ermöglichung einer Kurvenfahrt. Die Implementierung des *RotationNavigators* stellt hierfür keine Methoden zur Verfügung. Gerade dabei wird ersichtlich, dass die Berechnung der aktuellen Position sehr gut funktioniert. Zwar werden Kreisbögen grundsätzlich zu kurz gefahren, jedoch kann die restliche Strecke nachgefahren werden, weil die aktuelle Position des Fahrzeugs sehr genau berechnet wird. Die Ursache für

diesen Fehler liegt in der Regelung, da ein Schwingen um die zuvor berechnete Ideallinie erfolgt.

## 4 Diskussion

Aufgrund der Einschränkungen, welche die eingesetzte Hardware mit sich bringt, ist es nicht möglich, eine Fahrfigur exakt zu fahren. Dazu wäre es notwendig, die Genauigkeit der Rotationssensoren und gleichzeitig die Anzahl an Motorleistungsstufen zu erhöhen. Dennoch kann ein Roboter mit lejON hinreichend präzise fahren.

Die Evaluation der lejON-API hat ergeben, dass die Navigation gegenüber der in LeJOS verfügbaren nicht nur etwas genauer und die Kurvenfahrt nun als Fahrfigur verfügbar ist, sondern dass die Odometrie an sich enorm verbessert werden konnte. Begründet ist dies durch die hohe Anzahl an einstellbaren Fahrzeugparametern, insbesondere für den PID-Regler. Dadurch "weiß" der Roboter nun fast exakt, wo er sich befindet. Er kann sich also bei entsprechender Programmierung selbst korrigieren. Diese Eigenschaft ist allerdings nicht direkt in lejON implementiert, weil die Java Virtual Machine durch LeJOS und lejON beinahe gänzlich erschöpft ist und nur wenig Speicherplatz verbleibt. Der Anwender soll selbst entscheiden können, wie wichtig ihm diese Eigenschaft ist oder ob er lieber anderen Funktionen höhere Priorität einräumt.

Während der Evaluation kam nur eine Fahrzeugkonfiguration zum Einsatz. Für eine genauere und allgemeine Betrachtung sollten alle möglichen Werte der Hardwareparameter – wie Radgröße, Übersetzungsverhältnisse und Reifenprofil – getestet werden. In diesem Zusammenhang muss auch deutlich gemacht werden, dass die lejON-API für verschiedene Antriebsarchitekturen geeignet ist. Während der Realisation konnte jedoch nur eine, nämlich das DD, entwickelt und genutzt werden. Für weitere Architekturen war die Entwicklungszeit zu kurz.

Die für dieses Projekt eingesetzte Hardware basiert auf dem Lego-Mindstorms-Baukasten "Robotic Invention System 2.0" aus dem Jahre 1998. Wie bereits erörtert, ist diese mittlerweile veralteten Technik in ihrer Anwendung deutlich begrenzt.

Die Lego Group hat jedoch im August 2006 eine neue Mindstorms-Generation namens NXT veröffentlicht (vgl. [McN06]). Neben vielen kleineren Verbesserungen ist insbesondere die Hardware erneuert worden. So können selbstentwickelte Programme wesentlich genauer agieren.

Laut der Entwickler des LeJOS-Projekts wird auch LeJOS selbst an die neue Generation angepasst werden (vgl. [And02]). Aus diesem Grund bietet es sich an, für NXT auch lejON bereitzustellen. Durch die effiziente Schichtenarchitektur ist es lediglich notwendig, die Low-Level-Funktionen anzupassen, wie etwa die Methoden zur Ansteuerung der Motoren oder zum Auslesen der Rotationssensoren. Die High-Level-Funktionen können hingegen unverändert bleiben. So ist es für den Anwender bequem und einfach, neue Programme mit Hilfe von lejON zu entwerfen. Wie in dem aktuellen API muss er lediglich für neue Konstruktionen auch neue PID-Konstanten bestimmen, die er dann in den Initialisierungsmethoden einsetzt.

## Literatur

- [And02] Andrews, P. et al.: *LeJOS, Java for the RCX*. URI: <http://lejos.sourceforge.net/>, Stand: 22.03.2006, 2002.
- [Bau00] Baums, David: *Daves Baums Lego Mindstorms Roboter*. Galileo Press, 1. Auflage, 2000.
- [CH05] Chen, Quiang und Martin Hirzer: *Odometrie, Wo bin ich?* URI: [http://www.igi.tugraz.at/lehre/RoboSem/SS05/Slides\\_Odometrie.pdf](http://www.igi.tugraz.at/lehre/RoboSem/SS05/Slides_Odometrie.pdf), Stand: 24.08.2006, 2005.
- [DJ00] Dudek, Gregory und Michael Jenkin: *Computational principles of mobile robotics*. Cambridge University Press, 2000.
- [GRRS06] Grohmann, Jan, Annedore Rößling, Florian Ruh und Kai Schories: *lejON – LeJOS Odometric Navigator, Projektdokumentation*. URI: <http://lejon.sourceforge.net>, Stand: 15.06.2006, 2006.
- [KBM98] Kortenkamp, David, Peter Bonasso und Robin Murphy: *Artificial intelligence and mobile robots: case studies of successful robot systems*. AAAI Press/ The MIT Press, 1998.
- [McN06] McNally, M.: *Lego Mindstorms NXT Key Product Features*. URI: <http://mindstorms.lego.com/press/2057/LEGO%20MINDSTORMS%20NXT%20Key%20Product%20Features.aspx>, Stand: 22.03.2006, 2006.



# Optimierung von mobilen Geosensornetzwerken unter Berücksichtigung sowohl der Phänomen- als auch Geosensornetzwerk-Charakteristika

Alexander Christoph Walkowski, Phone: ++49.251-83-30056, Fax: ++49.251-83-39763, E-Mail: walkowski@uni-muenster.de, Institut für Geoinformatik der Westfälischen Wilhelms Universität zu Münster, Robert-Koch-Str. 26-28, 48149 Münster

## Abstract

In dem vorliegenden Beitrag wird ein Verfahren zur Messnetzoptimierung vorgestellt, welches die Charakteristik des räumlichen Phänomens aus den Messwerten schätzt und als Grundlage für die Optimierung nutzt. Basierend auf dem Phänomenmodell wird der Bedarf an zusätzlicher Information ermittelt. Bei der Optimierung des Netzwerkes werden die Einschränkungen des Geosensornetzwerkes explizit berücksichtigt.

**Keywords:** Geosensornetzwerk, Messnetzoptimierung, raumzeitliche Modellierung

## 1 Einleitung

Zur Beobachtung natürlicher, aber auch anthropogen bedingter Umweltphänomene ist die Erfassung der observierten Phänomene mit Hilfe von Sensoren unerlässlich. Erst die Sensorgestützte Erfassung macht die Phänomene für den Menschen sicht-/fassbar. Die *zeitliche Dimension* eines Phänomens beschreibt wie sich ein Phänomen an einem fixen Ort über die Zeit verhält. Zur Analyse werden Zeitreihen einzelner Sensoren betrachtet. Da Umweltphänomene – wie beispielsweise die Lufttemperatur oder die Konzentration bestimmter Luftschadstoffe – nicht punktuell auftreten, sondern am Besten durch kontinuierliche Werteoberflächen im Raum beschrieben werden können, bedarf es der Berücksichtigung einer weiteren Dimension: Raum. Im Hinblick auf die beispielhaft genannten Phänomene kommen zur Zeit nur *in-situ*<sup>1</sup> Sensoren zum Einsatz, die Messwerte für den Ort liefern, an dem sie installiert sind. Zur Erfassung der *räumlichen Dimension* eines Phänomens erfordert es die Installation von Sensoren in Form eines über den Raum verteilten Messnetzes. Exemplarisch sei hier das Luftqualitäts-Überwachungs-System (LUQS) des Landes-Umweltamtes Nordrhein-Westfalen (LUA-NRW 2001) genannt. Ziel dieses – primär aus stationären in-situ Sensoren aufgebauten – Netzwerkes ist es, Parameter der Luftqualität (Feinstaubbelastung, Stickstoffdioxid, Ozonkonzentration, etc.) zu erfassen.

*Traditionelle Verfahren* zur Datenerfassung für das Monitoring räumlicher Phänomene, wie das skizzierte LUQS-Messnetz, basieren auf einer geringen Anzahl von fixen Sensoren (SZEWCZYK et al. 2004), die in einer kontrollierten Umgebung mittels einer vorher festgelegten Deployment-Strategie installiert werden. Mit Hilfe der Strategie werden zum Beispiel die Position des Sensors und die notwendigen Kalibrierungsparameter ermittelt.

Durch die fortschreitende Entwicklung im Bereich der Sensortechnologie revolutioniert sich die Art und Weise, in der Daten über raumzeitliche Phänomene erfasst werden. Durch die

---

<sup>1</sup> Neben in-situ Sensoren gibt es auch noch Fernerkundungssensoren (remote Sensoren), die Charakteristika von entfernten Objekten erfassen, ohne dass sie in direktem Kontakt mit dem Objekt stehen.

Weiterentwicklung der Sensorik insbesondere der Evolution der micro-electro-mechanical-systems (MEMS) wandelt sich derzeit die Erfassung räumlicher Phänomene von einem auf isolierten Sensoren basierenden, zentralisierten Ansatz zu einem *Ansatz mit verteilten Sensoren* (NITTLE & STEFANIDIS 2005). Die Frage der Messnetzoptimierung ist für ein effektives Monitoring von großer Bedeutung.

Gegenstand aktueller Forschungsarbeiten ist – parallel zur Hardware-Entwicklung – der Aufbau von Kommunikationsinfrastrukturen, die an die Anforderungen in Sensornetzwerken angepasst sind (NITTLE et al. 2004) und die Integration einer Vielzahl von Sensorknoten in ein Sensornetzwerk ermöglichen. Exemplarisch seien hier die Entwicklung von speziellen Netzwerkprotokollen (GOLDIN et al. 2005 und WANG & RAMANATHAN 2004), Konzepte zur Selbstorganisation (BROOKS 2004) und Mobilitätsstrategien (SLIWINSKI & SIMONIS 2005) genannt.

In dem vorliegenden Beitrag wird ein Verfahren zur Messnetzoptimierung vorgestellt, welches die Charakteristik des räumlichen Phänomens aus den Messwerten schätzt und als Grundlage für die Optimierung nutzt. Basierend auf dem Phänomenmodell wird der Bedarf an zusätzlicher Information ermittelt. Bei der Optimierung des Netzwerkes werden die Einschränkungen des Geosensornetzwerkes explizit berücksichtigt.

Die vorliegende Arbeit gliedert sich wie folgt: Zunächst werden die zentralen Begriffe definiert und die Voraussetzungen für das vorgestellte Verfahren erläutert. Im dritten Abschnitt werden bestehende Verfahren zur Messnetzoptimierung skizziert. Der kombinierte Ansatz zur Optimierung von Geosensornetzwerken wird in Abschnitt vier vorgestellt. Der Beitrag schließt mit einem Ausblick auf weitere Forschungsarbeiten.

## 2 Randbedingungen

In diesem Abschnitt werden die für den Rahmen dieser Arbeit zentralen Begriffe definiert. Im Anschluss an die Definition eines Phänomens wird erläutert was unter einem mobilen Geosensornetzwerk zu verstehen ist. Da Geosensornetzwerke in der Regel nur punkthafte Messungen des kontinuierlichen Phänomens liefern, ist ein Phänomenmodell erforderlich, um eine kontinuierliche Repräsentation (Werteoberfläche) des Phänomens zu erhalten.

### 2.1 Phänomen

In dieser Arbeit wird ein *Phänomen* als ein raumzeitlicher stochastischer Prozess aufgefasst, der über eine räumliche Erhaltensneigung verfügt. Als räumliche Erhaltensneigung (*Persistenz*) wird das überzufällige Auftreten von ähnlichen Werten in räumlichen Nachbarschaften bezeichnet. Phänomene mit räumlicher Erhaltensneigung weisen an nahe beieinander gelegenen Orten ähnliche Werte auf. Zusätzlich kann das Phänomen über eine zeitliche Persistenz verfügen.

In Anlehnung an CRESSIE 1993 kann ein *raumzeitlicher stochastischer Prozess* als Menge  $Z$  von Zufallsvariablen  $Z(s, t)$  definiert werden,

$Z = \{Z(s, t), s \in D(t), t \in T\}$ , wobei  $s$  die Lokalisation der Zufallsvariablen hinsichtlich der räumlichen und  $t$  hinsichtlich der zeitlichen Dimension angibt.

Die Messwerte  $z(s,t)$  werden als Realisationen der Zufallsvariablen  $Z(s,t)$  angenommen; die Erfassung der Realisationen erfolgt durch ein mobiles Geosensornetzwerk.

## 2.2 Mobiles Geosensornetzwerk

Als *Geosensornetzwerk* (GSN) werden verteilte Sensornetzwerke bezeichnet, die auf die Erfassung raumzeitlicher Phänomene spezialisiert sind (STEFANIDIS & NITTLE 2005). Ein verteiltes Sensornetzwerk (distributed sensor network: DSN) besteht aus einer großen Anzahl heterogener intelligenter Sensoren, die über den Raum verteilt sind, über ein Kommunikationsnetzwerk miteinander in Verbindung stehen und auftretende Ereignisse an den Nutzer melden (IYENGAR et al. 2004). Wenn die Knoten (Sensoren) dieses Netzwerkes sich aktiv oder passiv im Raum bewegen können, spricht man von einem *mobilen Geosensornetzwerk*. Diese können ihr Verhalten basierend auf Stimuli – Beobachtungen, Nachrichten von Nutzern oder Komponenten zur Messnetzoptimierung – anpassen. Die Sensoren unterliegen Einschränkungen hinsichtlich der verfügbaren Energie, Mobilität (Bewegungsgeschwindigkeit, -radius).

Die Anwendung solcher GSN stellt ein Instrument dar, welches uns einen zuvor ungeahnten Einblick in raumzeitliche Phänomene erlaubt (siehe z.B. SZEWCZYK et al. 2004 und SHEPHERD & KUMAR 2004).

## 2.3 Zentrales Phänomenmodell

Die durch das mobile GSN erfassten Realisationen des raumzeitlichen stochastischen Prozesses stellen im statistischen Sinne die Stichprobe dar. Diese hinsichtlich der räumlichen und zeitlichen Dimension endliche Stichprobe ist keine adäquate Repräsentation eines kontinuierlichen Phänomens. Im Rahmen dieser Arbeit wird davon ausgegangen, dass alle Messwerte der Sensoren an eine Basisstation propagiert werden und dort ein Modell des beobachteten Phänomens erstellt wird.

Bei der Generierung des Modells werden aus den punktbezogenen Messwerten Realisationen der Zufallsvariablen an unbeprobten Orten geschätzt. Für die räumliche Interpolation wird das Kriging Verfahren (MATHERON 1963) genutzt, welches im Sinne eines autoprojektiven Verfahrens (STREIT 1981) nur auf der Grundlage der Messwerte eine im statistischen Sinne optimale Interpolation ermöglicht. Die räumliche Erhaltensneigung des räumlich stochastischen Prozesses wird beim Kriging durch das Variogramm modelliert.

Die Varianz des Schätzfehlers definiert die *Kriging-Varianz*

$\sigma_k = \text{VAR}(Z(s_0) - \hat{Z}(s_0))$ , wobei  $Z(s_0)$  der wahre Wert an  $s_0$  und  $\hat{Z}(s_0)$  der Kringingschätzer für  $s_0$  ist.

Die Kriging-Varianz ist abhängig von der Anzahl der für die Schätzung benutzen Stützstellen, dem Ort der Stützstellen und vom Variogramm, aber nicht von den Messwerten der Stützstellen (CRESSIE 1993:315). Sie beschreibt die durch die Messnetzkonfiguration entstehende Schätzunsicherheit und kann somit als Maß für fehlende Information genutzt werden. In Bereichen mit einer hohen Informationsdichte nimmt die Kriging-Varianz minimale Werte an.

Das zentrale Phänomenmodell liefert somit für den Nutzer eine adäquate Repräsentation des beobachteten Phänomens und für die Optimierung des Messnetzes ein Maß für fehlende Information.

## 2.4 Szenario

Es wird folgendes Szenario angenommen: In einem Untersuchungsgebiet ist ein Geosensornetzwerk zur Überwachung der Luftqualität installiert. Das GSN besteht aus zwei Subnetzen: (a) einem traditionellen stationären Messnetz ( $GSN_s$ ) und (b) einem mobilen GSN ( $GSN_m$ ). Nutzer des  $GSN_m$  können raumzeitliche Wegpunkte definieren, d.h. ein Sensor soll an einem Zeitpunkt oder für eine Zeitspanne Messungen an bestimmten Positionen im Raum vornehmen.

Bei einem Chemieunfall tritt eine Schadstoffwolke aus. Die Schadstoffkonzentration soll durch das GSN optimal erfasst werden. Da sich die Schadstoffwolke mit der Zeit ausbreitet, muss das  $GSN_m$  optimiert werden, um möglichst viel Information über das raumzeitliche Phänomen zu erfassen.

## 3 Messnetzoptimierung

Ziel der Messnetzoptimierung soll es sein, die Messnetzkonfiguration so anzupassen, dass die Sensoren effektiv eingesetzt werden. D.h. Sensoren sollen dort messen, wo sie im Modell noch nicht bekannte Information sammeln. Des Weiteren soll ein Modell des Phänomens errechnet werden, das einen minimalen Mangel an Information aufweist. Im Folgenden werden zwei Gruppen von Verfahren zur Messnetzoptimierung skizziert.

### 3.1 Coverage orientierte Ansätze

Die Abdeckung (coverage) eines Geosensornetzwerkes beschreibt einen Qualitätsparameter, der angibt, wie gut ein Untersuchungsgebiet von Sensoren abgedeckt wird (MEGUERDICHIAN et al. 2001, HOWARD et al. 2002). Schwachstellen in der Abdeckung bedeuten, dass Ereignisse im Untersuchungsgebiet auftreten können, jedoch nicht durch das GSN erfasst werden.

ZOU & CHAKRABARTY 2003 stellen beispielsweise einen sog. Virtual Force Algorithmus (VFA) vor, der in einer einmaligen Repositionierungsphase nach einer zufälligen Verteilung der Sensoren eine optimale Abdeckung gewährleistet. Sensoren, die gemäß einem vorher definierten Grenzwert zu nah beieinander liegen, stoßen sich ab, um zu starke Clusterbildung zu vermeiden. Hindernisse wirken ebenfalls abstoßend, um Kollisionen zu umgehen. Zu weit von einander entfernte Sensoren ziehen sich an, um so zu einer möglichst gleichmäßigen Verteilung der Sensoren im Untersuchungsgebiet zu gelangen.

Im Wesentlichen basieren coverage orientierte Ansätze auf geometrischen Überlegungen. Defizit dieses Ansatzes ist, dass die Messwerte nicht bei der Optimierung berücksichtigt werden und somit keine an das Phänomen sondern an die Geometrie des Untersuchungsgebietes angepasste Lösung erzielt wird.

### 3.2 Phänomen orientierte Ansätze

Im Gegensatz zu den coverage orientierten Ansätzen, bei denen die aktuelle Ausprägung des Phänomens nicht berücksichtigt wird, erfolgt bei Phänomen orientierten Ansätzen die Optimierung auf der Grundlage der aktuellen Messwerte.

SLIWINSKI & SIMONIS 2005 beschreiben unter anderem eine autonome Sensorstrategie, bei welcher ein Sensor solange an einem Ort verweilt wie er das Phänomen dort erfasst.

Andernfalls bewegt er sich zufällig durch den Raum. Bei diesem Ansatz wird das Phänomen explizit berücksichtigt; eine ganzheitliche Optimierung des GSN erfolgt jedoch nicht.

In PARDO-IGÚZQUIZA & DOWD 2005 wird ein Verfahren vorgestellt, welches auf der Grundlage der interpolierten Werteoberfläche und den assoziierten Schätzfehlern basiert. Der Ansatz nutzt die Kriging-Varianz, um schrittweise einen Sensor zu positionieren, sodass die räumliche Ausdehnung einer Wolke optimal erfasst wird. Für GSN ist die Annahme eines einzelnen Sensors unrealistisch. Nachteil des Ansatzes ist, dass nicht klar ist, welcher Sensor sich zu welchem Gebiet mit hohem Informationsmangel (hohe Kriging-Varianz) bewegen soll.

## 4 Phänomen- und GSN-basierter Ansatz

Die im vorherigen Abschnitt vorgestellten Verfahren zur Messnetzoptimierung eignen sich nur bedingt für die Optimierung eines GSN. Bei den coverage orientierten Ansätzen wird die Realisation des Phänomens nicht berücksichtigt. Die Messnetze werden nicht an die Realisation des dynamischen Phänomens, wie beispielsweise die Ausbreitung von Schadstoffwolken, angepasst. Die Erfassung entsprechender Phänomene kann somit nicht effektiv erfolgen.

Der im Folgenden vorgestellte Ansatz kombiniert ein Phänomenmodell mit einem Modell des GSN, um das Messnetz zu optimieren.

### 4.1 Phänomenmodell

Mit Hilfe des Kriging-Verfahrens wird aus der durch das GSN erfassten Stichprobe des Phänomens ein stochastisches Modell des Phänomens generiert. Wie oben dargestellt kann somit für den Nutzer des GSN eine adäquate Repräsentation des beobachteten Phänomens erstellt werden. Des Weiteren kann die Kriging-Varianz als Maß für fehlende Information verwendet werden.

Aus den Messwerten des GSN für den Zeitpunkt  $t_i$  ( $z(t_i)$ ) wird ein Modell generiert und die Kriging-Varianz berechnet, die angibt wie hoch der Informationsmangel zum Zeitpunkt  $t_i$  an jedem Ort im Untersuchungsgebiet ist. Erfolgte auf dieser Basis die Optimierung des GSN, bliebe die Dynamik des Phänomens unberücksichtigt. Die Optimierung würde aufgrund des Informationsmangels zum Zeitpunkt  $t_i$  berechnet, die nächste Messung erfolgt jedoch zu  $t_{i+1}$ . Zwischen  $t_i$  und  $t_{i+1}$  hat sich jedoch auch das raumzeitliche Phänomen verändert. Intuitiv würde man zu Verfahren der Zeitreihenanalyse greifen, um die zeitliche Variabilität zu berechnen und in die Optimierung einfließen zu lassen. Im Zusammenhang mit mobilen GSN sind keine langen Zeitreihen gegeben, sondern häufig nur einzelne Messwerte. Stattdessen wird nicht die Kriging-Varianz von  $t_i$  in die Optimierung integriert, sondern das Maximum der letzten  $n$  Kriging-Varianz Felder:

$$\sigma_{k,\max}(n) = \max(\sigma_k(t_i), \sigma_k(t_{i-1}), \dots, \sigma_k(t_{i-n}))$$

Somit wird errechnet, wo der größte Informationsmangel in den letzten  $n$  Zeitschritten bestand. In dem Feld  $\sigma_{k,\max}(n)$  werden Gebiete mit hohem Informationsmangel (Maxima) identifiziert. Im zweiten Schritt werden die Sensoren ermittelt, die in der Lage sind, das Zentrum dieser Regionen mit hohem Informationsmangel zu erreichen. Die triviale Lösung, den nächstgelegenen Sensor zu entsenden, schlägt bei heterogenen GSN, in dem sich nicht alle

Knoten bewegen können, fehl. Des Weiteren können keine *raumzeitlichen Wegpunkte* berücksichtigt werden, das heißt, es kann nicht berücksichtigt werden, dass ein mobiler Sensor zu bestimmten Zeiten an definierten Positionen Messungen vornehmen muss. Durch raumzeitliche Wegpunkte schränkt sich der Aktionsraum eines mobilen Sensors ein und die Bestimmung des Sensors, der in der Lage ist, das Zentrum der Region mit hohem Informationsmangel zu erreichen, erschwert sich.

## 4.2 GSN-Modell

Das Modell des GSN enthält sowohl eine Beschreibung des aktuellen Zustandes des GSN als auch eine Charakterisierung des potentiellen Aktionsraumes eines jeden Sensors. Hierzu wird auf das Konzept der Zeitgeographie von HÄGERSTRAND 1970 zurückgegriffen. Das Mobilitätspotential (Bewegungsgeschwindigkeit, Reichweite, etc.) jeden Sensors wird bei der Konstruktion eines Raum-Zeit-Prismas für einen Sensor berücksichtigt. Das Raum-Zeit-Prisma beschreibt den Aktionsraum eines Sensors, d.h. die Fläche, die von einem Sensor erreichbar ist, wenn er zu  $t_i$  an Position  $s(t_i)$  verweilt. Hierbei können auch raumzeitliche Wegpunkte berücksichtigt werden. Wenn ein Sensor zum Zeitpunkt  $t_{i+5}$  an einer bestimmten Position  $s(t_{i+5})$  eine Messung durchführen soll, verringert sich der Aktivitätsraum für den Zeitraum  $\Delta t = t_{i+5} - t_i$ . Ist kein Wegpunkt definiert, wird aus dem Raum-Zeit-Prisma ein Kegel, dessen Spitze in  $t_i$  liegt und sich in Richtung der Zeit öffnet. Die Grundfläche des Kegels beschreibt den potentiellen Aktionsraum. Für die Knoten des GSN<sub>s</sub> beträgt die Fläche des Aktionsraumes logischerweise null.

Unter Verwendung der Konzepte der Zeitgeographie können die Sensoren ermittelt werden, die potentiell das Zentrum einer Region mit hohem Informationsmangel ( $R_{IM}$ ) erreichen können. Dies ist genau dann der Fall, wenn das Zentrum von  $R_{IM}$  im Raum-Zeit-Prisma enthalten ist. Für den Fall, dass ein Zentrum in den Raum-Zeit-Prismen mehrerer Sensoren enthalten ist, wird der Sensor ausgewählt, der den kürzesten Weg zurücklegen muss. Somit ist gewährleistet, dass das GSN sich mit minimalem Energieverbrauch optimal organisiert.

## 5 Fazit und Ausblick

Das vorgestellte Verfahren zur Optimierung mobiler GSN ermöglicht die explizite Berücksichtigung sowohl der Charakteristika des Phänomens als auch des Geosensornetzwerkes. Der beschriebene Ansatz basiert auf der Minimierung der Kriging-Varianz als Maß für den Informationsmangel. Durch die Modellierung von Eigenschaften der Sensorknoten mittels der Zeitgeographie werden die Einschränkungen der Sensoren bei der Optimierung berücksichtigt.

Zurzeit basiert das Verfahren auf theoretischen Überlegungen. Gegenstand zukünftiger Arbeiten wird es sein, den vorgestellten Ansatz zu realisieren und auf seine Praxistauglichkeit hin zu prüfen. Im Rahmen der genannten Folgearbeiten muss auch die Auswirkung der zeitlichen Aggregation bei der Berechnung von  $\sigma_{k,\max}(n)$  analysiert werden.

## Literatur

- BROOKS, R. R. (2004): Need for Self-Configuration. In: IYENGAR, S. S. & R. R. BROOKS [Hg.] (2004): Distributed Sensor Networks. Boca Raton, 847-854. (=Computer and Information Science Series, 2)
- CRESSIE, N. A. C. (<sup>revised edition</sup>1993): Statistics for spatial data. New York. (=Wiley series in probability and mathematical statistics, .)
- GOLDIN, D., M. SONG, A. KUTLU, H. GAO & H. DAVE (2005): Georouting and Delta-Gathering: Efficient Data Propagation Techniques for GeoSensor Networks. In: STEFANIDIS, A. & S. NITTLE [Hg.] (2005): GeoSensor Networks. Boca Raton, 73-95.
- HÄGERSTRAND, T. (1970): What About People in Regional Science? In: Papers of the regional science association, 24: 6-21.
- HOWARD, A., M. J. MATARIC & G. S. SUKHATME (2002): Mobile Sensor Network Deployment using Potential Fields: a Distributed, Scalable Solution to the Area Coverage Problem. Proceedings of: 6th International Symposium on Distributed Autonomous Robotics Systems, June 25-27, 2002, Fukuoka, 299-308.
- IYENGAR, S. S., A. TANDON, Q. WU, E. CHO, N. S. V. RAO & V. K. VAISHNAVI (2004): Deployment of Sensors: An Overview. In: IYENGAR, S. S. & R. R. BROOKS [Hg.] (2004): Distributed Sensor Networks. Boca Raton, 483-504. (=Computer and Information Science Series, 2)
- LUA-NRW [Hg.] (2001): Luftqualität in Nordrhein-Westfalen. LUQS-Jahresbericht 1999. Essen.
- MATHERON, G. (1963): Principles of Geostatistics. In: Economic Geology, 58: 1246-1266.
- MEGUERDICHIAN, S., F. KOUSHANFAR, M. POTKONJAK & M. B. SRIVASTAVA (2001): Coverage Problems in Wireless Ad-hoc Sensor Networks. Proceedings of: InfoCom 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies, April 22-26, 2001, Anchorage, 1380-1387.
- NITTLE, S., M. DUCKHAM & L. KULIK (2004): Information Dissemination in Mobile Ad-Hoc Geosensor Networks. In: EGENHOFER, J. M., C. FREKSA & H. J. MILLER [Hg.] (2004): Geographic Information Science - Third International Conference, GIScience 2004 Adelphi, MD, USA, October 2004 Proceedings. Berlin and Heidelberg, 206-222. (=LNCS, 3234)
- NITTLE, S. & A. STEFANIDIS (2005): GeoSensor Networks and Virtual GeoReality. In: STEFANIDIS, A. & S. NITTLE [Hg.] (2005): GeoSensor Networks. Boca Raton, 1-9.
- PARDO-IGÚZQUIZA, E. & P. A. DOWD (2005): Multiple indicator cokriging with application to optimal sampling for environmental monitoring. In: Computers & Geosciences, 31(1): 1-13.
- SHEPHERD, D. & S. KUMAR (2004): Microsensor Applications. In: IYENGAR, S. S. & R. R. BROOKS [Hg.] (2004): Distributed Sensor Networks. Boca Raton, 11-27. (=Computer and Information Science Series, 2)
- SLIWINSKI, A. & I. SIMONIS (2005): An Experiment on Geosensor Mobility Strategies in the Planar Space. Proceedings of: 8th AGILE Conference on GIScience, May 26-28 2005, Estoril, Portugal, 177-186.
- STEFANIDIS, A. & S. NITTLE [Hg.] (2005): GeoSensor Networks. Boca Raton.

A.Ch. Walkowski

- STREIT, U. (1981): Kriging - eine geostatistische Methode zur räumlichen Interpolation hydrologischer Informationen. In: *Wasserwirtschaft*, 71(7/8): 219-223.
- SZEWCZYK, R., E. OSTERWEIL, J. POLASTRE, M. HAMILTON, A. MAINWARING & D. ESTRIN (2004): Habitat Monitoring with Sensor Networks. In: *Communications of the ACM*, 47(6): 34-40.
- WANG, K.-C. & P. RAMANATHAN (2004): Location-Centric Networking in Distributed Sensor Networks. In: IYENGAR, S. S. & R. R. BROOKS [Hg.] (2004): *Distributed Sensor Networks*. Boca Raton, 555-571. (=Computer and Information Science Series, 2)
- ZOU, Y. & K. CHAKRABARTY (2003): Sensor Deployment and Target Localization Based on Virtual Forces. *Proceedings of: InfoCom 2003, Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies*. March 30-April 3, 2003, San Francisco, 1293-1303.



# Anwendungsentwicklung und Praxistests eines mobilen Geodatenerfassungssystems für einen tragbaren Xybernaut-Computer

Dipl.-Inf. (FH) Frank Rudert, Phone: ++49.3943-659-361, Fax: ++49.3943-659-399, E-Mail: frudert@hs-harz.de , Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

Dipl.-Inf. (FH) Matthias Kulbe, mkulbe@yahoo.de

Prof. Dr. Hardy Pundt, Phone: ++49.3943-659-336, Fax: ++49.3943-659-399, E-Mail: hpundt@hs-harz.de , Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode

## Abstract

Die fortschreitende Miniaturisierung in der Elektrotechnik und der weiter wachsende Markt an mobilen Computersystemen eröffnen insbesondere im Bereich der Geoinformationssysteme neue Möglichkeiten zur Erfassung von Geodaten. Der Anwender kann beispielsweise direkt Daten vor Ort digital aufnehmen. In Verbindung mit GPS (bzw. in Zukunft Galileo) ist es möglich, neu erfasste Geoobjekte unmittelbar bezüglich ihrer Lage und ihrer Form zu beschreiben. Tragbare Computersysteme werden insbesondere im Hinblick auf derartige Anwendungsszenarien hin entwickelt.

## 1 Einleitung

Der Anspruch einiger Industriezweige hinsichtlich mobiler und flexibel einsetzbarer Computersysteme hat den Markt für so genannte wearable (anziehbare) Computersysteme geöffnet. Derartige Systeme müssen bestimmte Grundanforderungen erfüllen. Sie sollten möglichst bequem am Körper getragen werden können und Nutzer bei der Ausführung ihrer Arbeit nicht behindern. Insbesondere das Gewicht spielt hierbei eine Rolle. Zusätzlich müssen die Systeme so konstruiert sein, dass der Anwender bei Bedarf beide Hände für andere Tätigkeiten frei hat und dadurch nicht zwingend an das Computersystem gebunden sein darf.

Durch die fortschreitende Miniaturisierung im Elektronikbereich werden tragbare Systeme zunehmend praktikabler und erfüllen mittlerweile die Anforderungen hinsichtlich Größe und Gewicht. Bei individuellen Systemen wie Pulsmessenden Armbanduhren oder Kommunikationssystemen verschiedener Art ist die Entwicklung so weit fortgeschritten, dass sie bereits Einzug in den Alltag genommen haben. Bei flexibleren Systemen, welche ein möglichst breites Anwendungsspektrum erfüllen sollen, steht die Entwicklung in vielen Punkten noch am Anfang. Ein Grund, warum nicht nur Forschungseinrichtungen sich mit dieser Problematik beschäftigen, sondern auch in Unternehmen, beispielsweise IBM, verstärkt in die Entwicklung solcher Systeme investiert wird.

## 2 Hardware

Das Xybernaut-System besteht aus mehreren Komponenten. Das Herzstück stellt das PC-System dar, welches technisch einem Notebook ähnelt, jedoch aus Platzgründen in einer noch kompakteren Bauform integriert wurde und mit 900g angenehm leicht ist. Neben dieser Recheneinheit gehören ein Akku-Modul, ein Touch-Display, eine Maus mit Trackball, eine kompakte Tastatur und ein Headset mit integriertem Head Mounted Display zur Ausstattung. Ein aufsteckbares digitales Adaptermodul ermöglicht den Einsatz unterschiedlicher Schnittstellen.



Abb.1: Xybernaut MA-TC mit mobilem, digitalem Adaptermodul

### 3 Software

Zum Betrieb der Hardware wird ein reichhaltiges Softwarepaket benötigt. In erster Linie bildet das Betriebssystem die Grundlage für ein produktives System. Der Einsatz eines vollwertigen Desktop-Betriebssystems (Windows 2000) bringt einige Vorteile mit sich. In erster Linie ist der Treibersupport für Hardware wesentlich besser als bei proprietären Alternativen. Zusätzlich ist die Verfügbarkeit der GIS-Software für solche Betriebssysteme gegeben. Die Installation von alternativen Betriebssystemen wie Linux ist allerdings ebenfalls möglich und wird vom Hersteller für die Distribution Suse 8.1 oder höher unterstützt.

Für den Betrieb per Spracherkennung ist zusätzliche Software von Drittanbietern nötig. Erst Windows Vista wird eine eigene Spracherkennungssoftware mitliefern. Allgemein wird bei Sprachsteuerungssoftware in sprecherunabhängige und sprecherabhängige Spracherkennung unterschieden. Hierbei zeichnet sich die sprecherunabhängige Spracherkennung dadurch aus, dass vor der Verwendung keine explizite Aufzeichnung der Sprachkommandos nötig ist. Jedoch ist hier die Anzahl der verfügbaren Kommandos stark begrenzt, da die Software nur diese erkennen kann. Das Diktieren eines freien Textes ist hier nicht möglich. Bei sprecherabhängigen Systemen ist der Benutzer gezwungen, vor dem Betrieb die Software auf seine persönliche Aussprache zu trainieren. Dadurch wird der erkennbare Wortschatz bedeutend vergrößert, ist jedoch immer noch begrenzt.

Zur Nutzung einer Spracherkennungssoftware wird lediglich ein Mikrofon vorausgesetzt. Dies ist in dem Xybernaut beiliegenden Headset bereits integriert. Die Software bildet IBM Via Voice von der Firma Scansoft.

Seitdem Tablet-PCs am Markt verfügbar sind, ist die Verbreitung von Handschrifterkennungssystemen für Windows-Systeme erheblich gestiegen. Windows XP bietet eine integrierte Software zur Erkennung von Eingaben über den Touchscreen. In unseren Tests wurde die Software PenOffice der Firma Phatware für die Erkennung der Handschrift eingesetzt.

Auch die Praxistests (s. Kap. 4) haben gezeigt, dass die Eingabe per Touchscreen mit Hilfe von PenOffice die wohl einfachste Art ist, auch längere Texte ohne eine konventionelle Tastatur in das System einzugeben. Dies ist etwa in Bemerkungsfeldern im Rahmen der Erfassung von

Geodaten erforderlich. Da hierbei keine zusätzliche Maus oder Tastatur von Nöten ist, kann diese Art der Dateneingabe, sowie die Bedienung selbst, die Arbeit in vielen unterschiedlichen Umfeldern sehr erleichtern.

## 4 Feldversuch

Um die Praxistauglichkeit des MA-TC mitsamt der beschriebenen Komponenten prüfen zu können, sollte ein praxisnaher Feldtest durchgeführt werden. Dabei sollten mit Hilfe des MA-TC unterschiedliche Geoobjekte mit Hilfe eines GPS-Empfängers erfasst (Geometriedaten) und über eine Geodatenerfassungssoftware in Form so genannter Sachdaten näher beschrieben werden. Die Sachdaten werden vom Anwender eingegeben (Stift- oder Spracheingabe). Zur Erfassung der Daten sollte die Anwendung Gispad der conterra GmbH (Münster) eingesetzt werden.

### 4.1 Vorbereitung

Der Versuch sollte primär mit Hilfe des Touchdisplays, der mobilen Tastatur und dem mobilen Trackball durchgeführt werden. Zusätzlich sollten aber auch die weiteren Steuerungsmöglichkeiten, wie Sprachsteuerung und Handschriftenerkennung und die grafische Ausgabe auf dem Head Mounted Display getestet werden.

Vorbereitend wurde die nötige Software installiert. Dies bezieht sich hauptsächlich auf die Installation von Gispad, einem GIS-Frontend zur Geodatenerfassung. Um eine spezielle Anwendung mittels Gispad entwickeln zu können, ist es notwendig, über einen Objektklasseneditor, der quasi Bestandteil von Gispad ist, einzelne Objekte zu definieren und ihre Eigenschaften festzulegen. Jedes Geoobjekt wird anhand von Attributen näher beschrieben. Damit diese Geoobjekte im Freien komfortabel aufgenommen werden können, werden Formulare für jede Objektklasse erstellt. Die Dateneingabe erfolgt dann über diese Formulare, die somit zunächst entwickelt werden müssen, bevor sie dann in die Anwendung integriert und seitens des mobilen Gerätes zur Verfügung gestellt werden können.



Abb. 2:  
Kompatibilitätsproblem  
zwischen MA-TC und GPS-  
Empfänger

Gispad bietet die Möglichkeit einen GPS-Empfänger zur Positionsbestimmung einzubinden. Der für unseren Test eingesetzte GPS-Empfänger auf CompactFlash-Basis stellte jedoch eine mechanische Inkompatibilität dar, da der Einschub für die PCMCIA-Peripherie extrem eng ist und somit den Einsatz verschiedener Geräte von vornherein blockiert. Die Lösung würden GPS-Empfänger auf USB- oder PCMCIA-Basis bringen, deren Anschaffung im Rahmen dieser Praxistests jedoch nicht durchführbar war.

### 4.2 Ergebnisse und Fazit

Durch das bereits beschriebene Problem mit dem GPS-Empfänger ist der Feldtest im Bezug auf die Erfassung von Daten nur teilweise den Erwartungen gerecht geworden, da die Komponenten zwar einzeln, jedoch nicht im Gesamtpaket testbar waren. Trotzdem konnten wichtige Erkenntnisse hinsichtlich der Praxistauglichkeit des MA-TC sowie die zugehörige Peripherie gewonnen werden.

Das Anlegen des Systems soll zwar durch eine spezielle Weste vereinfacht werden, dies erwies sich jedoch ohne Hilfe einer anderen Person als schwerfällige Aufgabe. Insbesondere das Unterbringen der Kabel, welche für den Anschluss der einzelnen Komponenten nötig sind, ist zeitaufwändig. Während der Arbeit können die Kabelstränge den Benutzer stören. Teilweise passierte es auch, dass durch eine ungünstige Bewegung Stecker aus dem Xybernaut heraus gezogen werden, da diese nicht mit dem Hauptgerät verschraubbar sind, ein Mangel, der angesichts der durch die Werbung hoch gesteckten Erwartungen unbedingt zu beheben ist.

Tests mit dem Touchscreen ergaben, dass das Display bei starker Sonneneinstrahlung ähnlich anfällig wie herkömmliche Laptop-Displays ist und demzufolge die Lesbarkeit teilweise eingeschränkt war. Im Schatten, bei Bewölkung und fehlender direkter Sonneneinstrahlung ergeben sich gleichwohl keine Probleme.

Das Head Mounted Display erwies sich mitunter als zu klein; um simultan Daten, die über den Rechner geliefert werden, und die reale Situation zu sehen, was für spezielle mobile Anwendungen sehr nützlich sein kann, erfordert das vergleichsweise kleine Display hohe Konzentration. Die Verwendung der Tastatur (am Arm befestigt) und der Maus erwies sich als gut handhabbar. Einschränkend muss lediglich gesagt werden, dass die Eingabe von Sonderzeichen recht umständlich ist.

Die Sprachsteuerung reagierte teilweise zu sensibel auf Umgebungsgeräusche wie Straßenlärm oder stärkeren Wind. Zusätzlich irritiert ein Störgeräusch des MA-TC selbst die Eingabe von Befehlen. In ruhigen Umgebungen ist die Sprachsteuerung dennoch zu bevorzugen, da der Anwender hierbei beide Hände frei zum arbeiten hat.

Ein echter Mangel trat beim Wechsel des Akkus auf. Der im Betriebssystem angezeigte Ladezustand der Batterie bezieht sich nur auf die Notbatterie im Gerät selbst, welche immer die volle Kapazität besitzt, solange die externe Batterie angeklemt ist. Sobald die externe Batterie ausfällt wird auf die interne Notbatterie umgeschaltet, das Display schaltet sich ab. Der Benutzer hat nun keine Möglichkeit mehr das Gerät korrekt herunter zu fahren und kann nur hoffen, dass es ordnungsgemäß in den Ruhezustand überführt wird und Datenverlust verhindert wurde. Ist dieser Zustand dem Nutzer vorher nicht bekannt, so kann dies im Rahmen einer zeit- und damit kostenintensiven Geodatenerfassungskampagne zu nicht akzeptabler Zweiterfassung der vorab nicht gesicherten Daten führen.

Die Arbeit mit Gispad ist insbesondere durch die frei erstellbaren Formulare in Verbindung mit der Handschrifterkennung sehr komfortabel. Das Eingeben von Texten gelingt problemlos. Der Benutzer hat somit letztlich auch unterwegs immer die Softwareausstattung dabei, welche er im Büro gewohnt ist. Bisherige Systeme auf Basis von PDAs sind meist im Funktionsumfang stark begrenzt. Es ist beispielsweise auch möglich, noch im Gelände neue Objektklassen mit Gispad anzulegen, die dann unmittelbar mit erfasst und in der Datenbank abgelegt werden können.

Abschließend kann gesagt werden, dass der MA-TC für den Einsatz im Feld unter den unterschiedlichen Bedingungen durchaus geeignet ist. Dies verdankt er nicht zuletzt seinem flexiblen Equipment und den damit verbundenen vielfältigen Möglichkeiten zur Eingabe bzw. Anzeige von Daten. Nicht nur bei der satellitengestützten Erfassung von Geodaten kann ein solches Gerät behilflich sein, auch Mixed-Reality-Anwendungen sind auf der Basis eines solchen System denkbar. Dabei sollte die verwendete Software immer an die speziellen Anforderungen, welche durch die Ein- bzw. Ausgabeperipherie gegeben sind, angepasst werden, um Nutzer im Feld im Sinne eines Medienbruchfreien Arbeitens wirkungsvoll zu unterstützen. Neuere Geräte sollten an den aktuellen Stand der Technik angepasst, die

beschriebenen Mängel natürlich behoben werden. Hierbei sei auf neuere Stromsparende Prozessortechnologien und Kurzstreckenfunktechnik wie Bluetooth zur Verminderung des Kabelaufwandes für die Peripherie verwiesen.

Angesichts des rasanten Anstiegs bei der Nutzung mobiler Geräte auch für komplexere Anwendungen werden den wearable Computern zukünftig mehr Aufgaben im Bereich der Erfassung und Verarbeitung von Geodaten zukommen. Die hier beschriebenen Praxistests haben wichtige Erkenntnisse zur Arbeit mit derartigen Systemen und ihrer Kombinierbarkeit mit spezieller GIS-Software und GPS-Empfängern erbracht, die beim Ausbau des Systems und beim Einsatz in Forschung und Lehre Berücksichtigung finden werden.

### **Quellen:**

- XYBERNAUT: Homepage des Herstellers des MA-TC.  
Online unter: <http://www.xybernavt.de> (abgerufen am 02.02.2005)
- MA-TC TECH: Technische Spezifikation zum MA-TC auf der Seite des Herstellers.  
Online unter: [http://www.xybernavt.de/produkte/engl/e\\_ixma4tc.html](http://www.xybernavt.de/produkte/engl/e_ixma4tc.html)  
(abgerufen am 02.02.2005)
- MIT WEARABLE: Wearable Computing Webseite vom MediaLab des MIT. Online  
unter: <http://www.media.mit.edu/wearables/> (abgerufen am  
06.02.2005)
- TZI Bremen: Studie zum Thema mobiler, tragbarer Computersysteme der des TZI  
der Universität Bremen.  
Online unter: <http://matrix.wearlab.de/studie/studie.html> (abgerufen  
am 06.02.2005)