Modelling production of personalized information services and their delivery on multiple distribution channels

Dipl.-Inform. Thomas Ritz

Fraunhofer IAO
Nobelstr. 12
D-70569 Stuttgart
Thomas.Ritz@iao.fhg.de

Abstract. Personalized information services have become a typical information product offered online on different media, but mainly the web. After a short definition of personalized information services we propose a modelling formalism, defining the building blocks for the production of personalized services and is applicable as “glue” between business processes modelling and software engineering. Within this paper we will instance this model for a very simple personalization method and show how the former elaborated model could be mapped on a three-tier software architecture. Finally it is shown that these architecture eases the personalized delivery on different output channels (f.i. HTML, WML and VoiceXML), based on one single content repository.

Introduction to Personalized Information Service

Need for personalization

Personalized delivery of content is gaining importance, which is reflected in ongoing research. A huge number of such services are not primarily designed as information product like a newspaper. Mostly the services are designed to cope with 1:1 marketing aspects [1, 2]. Nevertheless the added value for publisher as well as for consumers of such information services is proved [3] and could be traded as part of value-added-publishing strategies [4]. As common subset of definitions which could be found in literature [5,6,7], we could state that personalized information services are service which deliver the right information at the right time on a media preferred by the consumer [8]. Information is considered in the following as multimedia information (text, audio and video) as proposed by Smyth [9]. Furthermore, we assume publishers in the right position to produce and deliver such services.

Definition of personalized information services

As stated before, lots of services are labelled personalized service. A definition based on common characteristics will result in a generic definition like

“Personalization is a service provided based on a user profile”. Which individual services generated using this profiles is still uncertain. The definition as follows (c.f. [3]) tries to integrate the dimensions found in definitions cited before:

- process
- user interest (textual)
- demand in schedule
- demand in media

Def. 1: A *Personalized Information Service* is a service towards a customer comprising

1. *filtering* of information out of former *gathered* and *qualified* information regarding users textual *interest*
2. *presentation* of this information using a user defined *time schedule* and *media* appropriate with recent user environment.

This brings us back a definition saying “personalized information services are services which deliver the right information at the right time on a media preferred by the consumer” [8].

Production of Personalized Information Services applying the mass customization paradigm

The production of information services

As for other products, the production process became part of competition in the media market [10]. Most of the efforts spent in research on production of media and information services focussed on “media independent publishing” strategies [11], which are quite often referred to as Cross-Media-Publishing ([12] or [13]). Hand in Hand with this developments it became standard to handle media elements in different formats and media during the production process as text, audio and video material [14]. To stick to this paradigm of cross-media publishing, markuplanguages as SGML and nowadays XML became common to handle and integrate the different media elements and to define a semantic structure for the information [15]. As the distribution media for information products lost focus in the production chain, more attention was spent to focus on market demands and target groups on information level [10]. The editorial work got more attention [14] but the discussion still sticks to the production of mass-media [16].

Modelling mass customized information services

We start setting up a theoretical framework by modelling the product to be produced. Same as for other customized or individual products, we assume modularity to be a crucucial factor for efficient production [17]. Therefore we assume a a building block structure [18] as shown in Fig. 1.

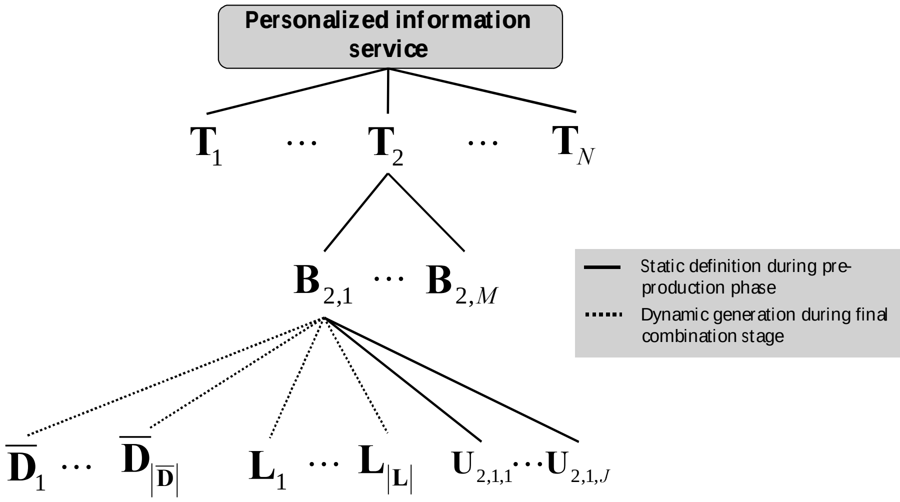


Fig. 1.: Building blocks for personalized informtion service

Depending on the user profile, the product shipped to the consumer consists of templates ($\mathbf{T}_i \in \mathbf{T}$) which aggregate building blocks ($\mathbf{B}_{i,j} \in \mathbf{B}$). $\mathbf{T}$ and $\mathbf{B}$ are sets of all available templates and building blocks. A building block represents a functional cluster (f.i. show teasers). So a building block may select a set of documents from the content repository ($\overline{\mathbf{D}}$) and renders it. Please note that this assignment of documents is not fixed. The doted line should indicate that the building block is dynamically populated with content, taking the user profile under consideration. Further to the documents, a building block may contain links ($\mathbf{L}$). Links are relations between instances of templates. Finally a building block may contain user interfaces for functions as login etc.. The example shown in Fig. 2 illustrates the construction of a sample home page with these building blocks.

Knowing about the structure helps to investigate the production process for personalized information services. We start with the personalization itself. It could be easily modelled starting with a generic information retrieval model as proposed e.g. by Fuhr [19] (for further formalisms see f.i. [20]). This model shown in Fig.3 shows that Documents D are in real world in certain relations R to user queries Q. In order to process them electronically, both real world entities have to be mapped to a digital representation (i.e. for vector space retrieval or an onthology based classification). This is done by the functions α . The resulting digital representations could be optmized by β at runtime to a representation appropriate for the particular information retrieval concept applied (i.e. handling of sparse matrices). Finally, the information retrieval is done by applying the information retrieval function $\Theta : \mathbf{D} \times \mathbf{Q} \rightarrow \mathbf{IR}$, which result in a rank of the documents with regard to a users query.

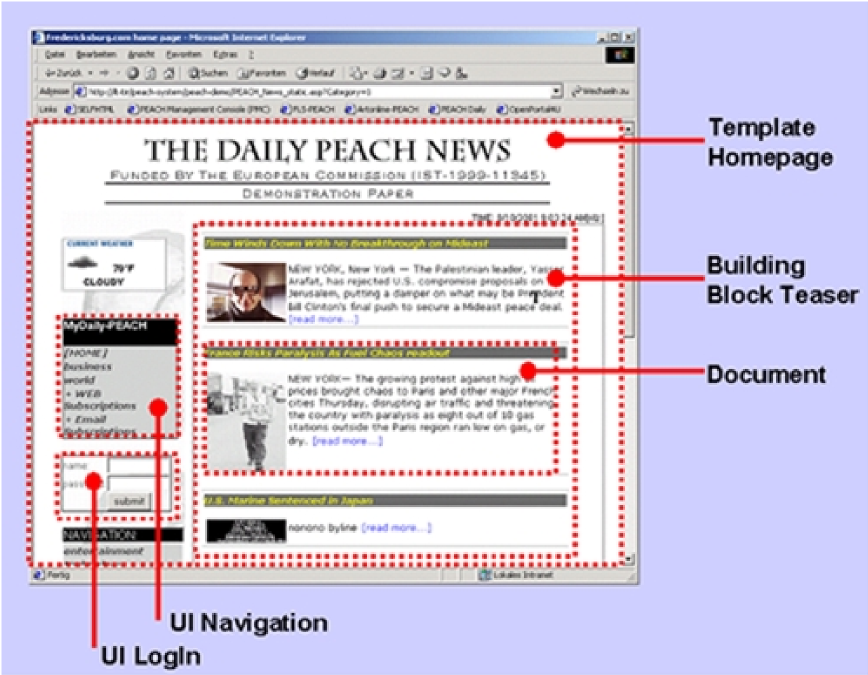


Fig. 2.: Sample of Buildingblock Structure

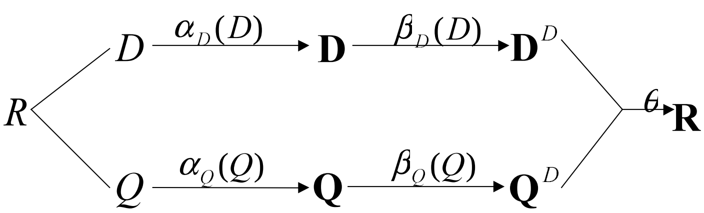


Fig. 3.: Information Retrieval Modell [19].

Personalization could be seen in fact as such an information retrieval process. The user query have to be seen as a set of user queries defining the users profile. But as seen, users preferences could contain much more than textual interest [3]. Therefore we define a user query as a data structure and a couple of methods to tailor this data structure:

The first component (C_p) of the vector is containing the description of preferences in contents (e.g. tennis). The second component (M_p) contains a media preference for this (i.e. WAP), while the last component (S_p) defines a schedule. So a user could profile his interest, e.g. for “business news every morning on the PDA”. To give the user the chance to tailor its profile, we introduced two functions to change the digital representation of the user profile. The first ($\tilde{\alpha}_Q(Q_P)$) is intended for manually changing a profile. The second one ($\hat{\alpha}_{Q_P}(\mathbf{D})$) is aimed to tailor the user profile with document representations (i.e. TF vectors). This can be used to tailor the user profile applying relevance feedback methodologies [20, 21].

$$Q_P = \begin{pmatrix} C_P \\ M_P \\ S_P \end{pmatrix} \xrightarrow[\tilde{\alpha}_Q(Q_P)]{\alpha_Q(Q_P)} \begin{pmatrix} \mathbf{C}_P \\ \mathbf{M}_P \\ \mathbf{S}_P \end{pmatrix} \xrightarrow[\hat{\alpha}_{Q_P}(\mathbf{D})]{\beta_Q(Q_P)} \begin{pmatrix} \mathbf{C}_P^D \\ \mathbf{M}_P^D \\ \mathbf{S}_P^D \end{pmatrix}$$

Fig. 4. Extended IR Modell for modelling user profiles

Further to this, the information retrieval model results in a ranking of contents. For personalized information services we have to result in an information product as modelled in Fig. 2.. So we extended the IR Modell again by the building blocks of a personalized information service (c.f. Fig. 1.) and resulting in a the model shown in Fig. 5.

In this model the user requests information from a certain media at a given schedule. The presentation function $\wp(P, M, S)$ renders the information product making use of the building block structure defined before. So a template appropriate with the requested media is called which contains several building blocks. This building blocks make use of the information retrieval function and retrieve the relevant documents (as in the standard IR Modell). While $\mathbf{D}$ is the digital representation for information retrieval (computed by α), the digital representation of the documents $\overline{\mathbf{D}}$ contains all material of an article elaborated during the editorial process. To tailor it for the special need in the building block, a filter $\mathbf{F}$ is employed by the function $\phi(\overline{\mathbf{D}}, \mathbf{F})$. This filters extract f.i. only the headline and the first paragraph for a teaser presentation (In fact $\overline{\mathbf{D}}$ and $\mathbf{F}$ are realized as XML and XSLT documents in the later described application).

For the editorial processing, we introduce the editor function $\varepsilon_D(D, Y)$, which is used to markup a document in the real word according to a given document type definition $\mathbf{Y}$, which defines the document structure. Due to the fact that $\mathbf{Y}$, $\mathbf{F}$, $\mathbf{B}$ and $\mathbf{T}$ are not static, we defined design functions δ for this purpose. The user interfaces to the user profile functions ($\alpha_Q(Q_P), \tilde{\alpha}_Q(Q_P), \hat{\alpha}_{Q_P}(\mathbf{D})$ and $\lambda(P)$) are collected in the set $\mathbf{U}_M$ for easy notation purpose.

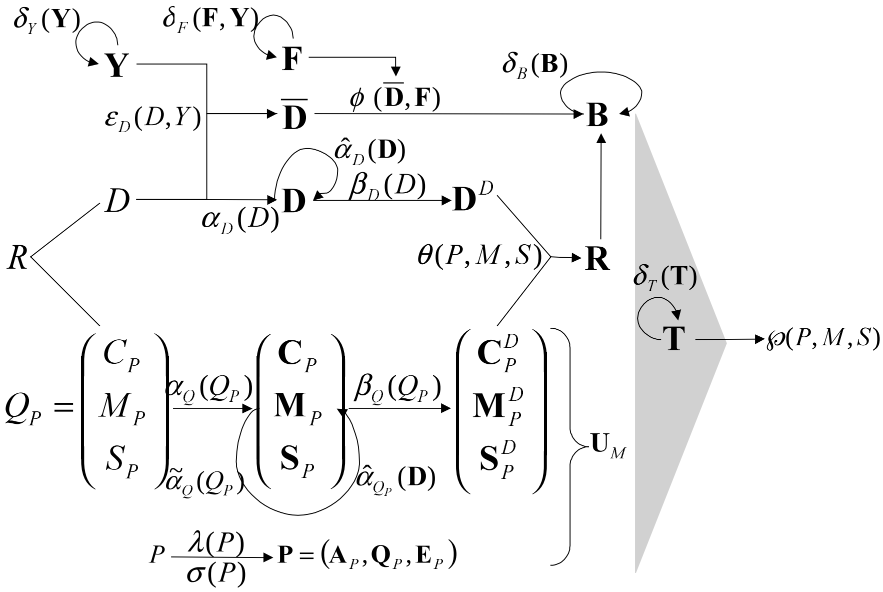


Fig. 5. Model of production and provision of personalized information services

Now we modelled all components needed to produce and provide personalized information services. The use of this model is manifold. We can use the model for requirements engineering by assigning model elements to production and workflow processes (c.f. [22]). Secondly, we can use the model for software engineering purpose by implementing the components. Our model eases this, cause dependencies of model components became transparent. But the greatest use is the mediation between requirements engineering and software engineering.

Employing the model

To show an instance of the former introduced formal model, we will employ the model for a very simple personalization application. The user checks on a multiple choice form, whether he is interested in certain categories or not. We assume that the contents were manually classified according the same classification scheme within the newsroom.

Let $\mathbf{KT} = (K, R)$ be a tree. K are the leaves of this tree and reflecting the categories. The edges $R = \{(k_i, k_j) \mid k_i, k_j \in K; 1 \leq i, j \leq |K|\}$ allow to order the leaves hirarchically. The manual assignments of documents to leaves form a relation $Z_D^M \subset K \times D$ between documents and leaves (vgl. [23]). This relations represents the digital representation of the Documents $\mathbf{D}$ and the mapping function α_D is the manual assignment. As mentioned before the user is asked to fill a multiple choice form in order to assign himself to parts of the tree $\mathbf{KT}$. Thus a relation $Z_Q^M \subset K \times Q$ is established and defines C_p, the component of the user profile describing the users textual interest. The retrieval function θ could now be easily computed by evaluating the product of the relations, thus $\theta = Z_Q \circ Z_D$. Table 1 shows the use of model components in a multi-tiered-architecture for the buiding block which is used to select and present teasers to the consumer.

	$b \in \mathbf{B}$	Building Block HTML Teaser
Presentation	$\forall d \in \theta :$ $\phi\ (d, f)\ \ f \in F$	 Nascar after attack ... Last week's terrorist attacks in New York and Washington will have an effect on the mood surrounding the MBNA Cal Ripken Jr. 400. And though the increased military presence in and around Dover Air Force Base will bring the reality of our world home that much more clearly, the direct impact on fans will be fairly minimal. That is, though there necessarily will be heightened security (see Question 2) throughout the weekend, few fans will bear any real consequences. [read more...] Clemens First Pitcher to Go 20-1 as Yankees Roll On Roger Clemens was not at his sharpest, striking out only one, in becoming the first pitcher to have a 20-1 record.The phone calls from family and friends began at noon today and came to Roger Clemens in rapid succession; good luck, best wishes. But no one mentioned how Clemens had the chance to become the first pitcher to win 20 of his first 21 decisions. They just kept reminding Clemens that he was a representative of New York. [read more...]
Logic	$\theta\ = Z_Q \circ Z_D$	Implemented as method, wich calls a SQL statement.
Data	$\overline{\mathbf{D}}\ \mathbf{D} \begin{pmatrix} \mathbf{C}_P \\ \mathbf{M}_P \\ \mathbf{S}_P \end{pmatrix} \mathbf{F}$	Relational DB ; $\overline{\mathbf{D}}$ XML documents; $\mathbf{F}$ XSLT documents

Table 1. : Definition of a building block

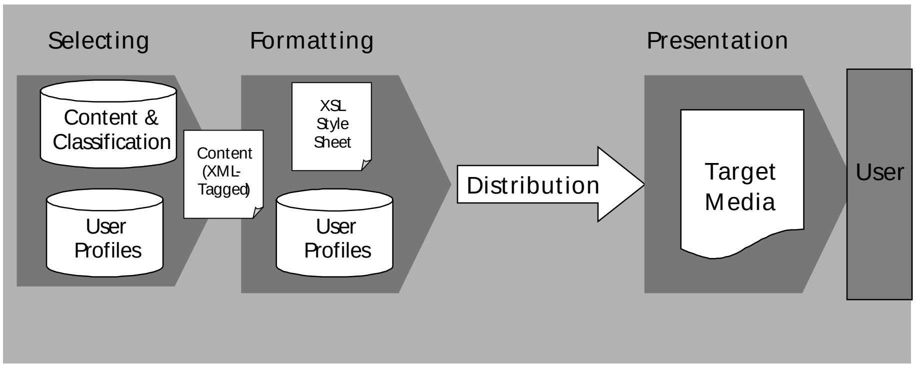
Within the European funded research project PEACH, several other Information Retrieval and classification methodologies were applied, as semantic classification and full text retrieval. The former presented model allowed us to compare these methods mutually. Further more we placed elements of the former mentioned personalization model in business process models (c.f. [22]). As a result you can see that f.i. manual classification causes bigger efforts in the editorial process and in the end limits the user to a small number of categories. In comparison automatic/semiautomatic classification causes rather no efforts for the editors, but is not easy to set up and to understand by the end users. Fulltext search causes again no additional efforts, the users are not mapped on a comparable classification scheme which makes recommendation and collaboration more difficult.

Implementations

As shown in the example the model could be easily transferred to a 3-tier software architecture [24, 25].

This approach ensures that the business logic components could be reused for several media, by handling the particular presentation in the presentation layer. The data layer comprises the atomic information. The business objects are designed to deal with basic functionality which can be easily embedded in the presentation layer and which encapsulates the data layer from direct access. This is extremely important for critical functions (i.e. dealing with the user profile), cause business rules i.e. for handling privacy issues have to be defined and validated once.

We realised these objects as COM objects which could be used by ASP scripts. The business objects fosters the programmer to the following process shown in Fig 6..



.

Fig. 6. The overall process of selecting and formatting contents could be shown as follows

First the information is selected by bringing user and content object together (this results in a set of contents appropriate for the recent user). Following to this the contents could be formatted within ASP pages (or any other COM capable scripting or programming language). Applying this principle, it became possible to easy generate different forms of personalized information (f.i. HTML and VoiceXML):

WWW

```
'SELECT

<%

Set CO =
CreateObject("ContentObject.ContentClass")

SET rs_contents_caption =
CO.Select_My_Categories(UO.UserID,"<web>")

%>

'FORMAT in HTML

<%

While Not qry_return.EOF

   response.write "<a
href='sample1.asp?Category=" &
qry_return("ID")& "'>" & qry_return("Name")
& "</a><br>"

   qry_return.MoveNext

Wend

%>
```

VoiceXML

```
'SELECT
<%

Set CO =
CreateObject("ContentObject.ContentClass")

SET rs_contents_caption =
CO.Select_My_Categories(UO.UserID,"<Voice>")

%>



'FORMAT in VoiceXML

<menu id="nyc_menu" dtmf="true">

  <prompt>

   <audio src="jingle.wav" />
```

```
    Hi, <%=UO.User%>

            My PEACH daily news have the
following categories for you

  </prompt>

  <enumerate>

    For <value expr="_prompt"/>, press
    <value expr="_dtmf"/>

  </enumerate>

<%

  do while not rs_contents_caption.EOF
    response.write "<choice
    next='http://lt-tir/peach-
    system/test/teaser_vxml.asp?
    category="& rs_contents_caption("ID")
    &"'>" &  rs_contents_caption("Name")

    response.write "<grammar>" &
    rs_contents_caption("Name") &
    "</grammar>" & "</choice>"

    rs_contents_caption.movenext

  loop

 %>
```

Table 2. Scripting of personalized navigation for HTML and VoiceXML browsers

Conclusion

After an introduction into personalized information services, the text presented a formal modelling method to describe personalized information services from an abstract and modular view. To prove its viability the model was applied for personalization based on categorized text's and user's. The model presented was mapped on a 3-tier architecture and a short introduction into the resulting object-oriented technology could be given. Finally the resulting business objects were shown "in action", to produce an personalized HTML and VoiceXML navigation based on a single content repository. In addition to the sample application refered to in this text, the model proved its viability within the PEACH project funded by the European Commission's IST programme (IST-1999-11345).

Dipl.-Inform. Thomas Ritz , born 1971 studied computer science and economics at the University of Bonn and obtained his degree in computer science in 1997. Since 1997 he is scientific employee at the University of Stuttgart IAT and the Fraunhofer IAO. He has a long standing experience in publishing industry and focuses his research activities on online communication and information. In several research projects he designed and implemented solutions for electronic information exchange, online communication in virtual communities, branch specific information portals and electronic commerce.

Mr. Ritz acts as project manager in the DISMED project (funded by the European Commission in FP 4 Innovation program RSE 067) and in the PEACH project (funded by the European Commission in FP 5 IST programme IST-1999-11345). Besides he joined several public and industry funded research, consultancy and development projects.

References:

[1] J.W. Palmer and L.B. Eriksen "Digital Newspapers Explore Marketing on the Internet," *Communications of the ACM* vol. 42, no. 9, pp. 33-40, 1999.

[2] H. Fisbeck "Customized Marketing im Internet," *Akademische Abhandlungen zu den Wirtschaftswissenschaften* 1999.

[3] T. Ritz "Personalized Information Services - An Electronic Information Commodoty," *e-Business Strategy Management* vol. 2, no. 2, pp. 105-114, 2000.

[4] H. Berghel "Value-Added Publishing," *Communications of the ACM* vol. 42, no. 1, pp. 19-23, 1999.

[5] C. Allen "Personalization vs. Customization," http://clickz.com/cgi-bin/gt/en/pm/pm.html, 2000.

[6] J. Nielsen "Personalization is Over-Rated," http://www.useit.com/alertbox/981004.html, 2000.

[7] IBM Developer Works "Web site personalization," http://www-4.ibm.com/software/developer/library/personalization/index.html, 2000.

[8] T. Ritz "Personalized Information Services - an electronic information commodity and its production," *ICCC/IFIF conference on Electronic Publishing (ELPUB 2001)* pp. 48-58, 2001.

[9] B. Smyth and P. Cotter "A Personalized Television Listing Service," *Communications of the ACM* vol. 43, no. 8, pp. 107-111, 2000.

[10] M. Freter "Crossmedia-Publishing - Vorbereitung aud den Wettbewerb von Morgen," *Jahrbuch 2000 der Fachinformationen* 2000.

[11] C. Shapiro and H.R. Varian "Information Rules," 1998.

[12] E. Fritz "Cross-Media: Ein Inhalt aber mehrere Darstellungsformen," *Deutscher Drucker* no. 7, pp. w31-w32, 2000.

[13] C. Fouchard "Eine ODBMS-basierte Lösung für "Cross-Media-Publishing"," *Objekt Spektrum* no. 3, pp. 91-95, 2000.

[14] B. Oakly, D. Kueter and K. O'Heas "The Future of Content," *Discussions on the future of European Electronic Publishing* 1997.

[15] H.-J. Bullinger, E. Schuster and S. Wilhelm "Content Management Systeme," 2000.
[16] M. Schumann and T. Hess "Grundlagen der Medienwirtschaft," 2000.
[17] J. Pine "Mass customizing products and services," *Plannung Review* vol. 21, no. 4, pp. 6-13, 1993.
[18] J. Göpfert and M. Steinbracher "Modulare Produktentwicklung leistet mehr," *Harvard Business Manager* no. 3, pp. 20-29, 2000.
[19] N. Fuhr "Information Retrieval - Skriptum zur Vorlesung," 1998.
[20] R. Baeza-Yates and B. Ribeiro-Neto "Modern Information Retrieval," 1999.
[21] C.v. Rijsbergen "Information Retrieval," 1979.
[22] T. Ritz "Modelling Mass Customization of digital news services," *e-Business Strategy Management* vol. 3, no. 3, 2002.
[23] W. Dörfler and W. Peschek "Einführung in die Mathematik für Informatiker," 1988.
[24] R. Peters "Business-Objekte, Workflow und die UML," *OBJEKTspektrum* vol. 3, pp. 69-73, 1999.
[25] J. Sturm "VB 6 UML Design and Development," 1999.