Issue 199

October 2015

The IWi Process Model Corpus

Tom Thaler, Sharam Dadashnia, Andreas Sonntag, Peter Fettke, Peter Loos

Publications of the Institute for Information Systems (IWi)
at the German Research Center for Artificial Intelligence (DFKI)

Editor: Prof. Dr. Peter Loos

T. THALER, S. DADASHNIA, A. SONNTAG, P. FETTKE, P. LOOS

The IWi Process Model Corpus

Publications of the Institute for Information Systems (IWi)
Editor: Prof. Dr. Peter Loos

Issue 199

ISSN 1438 5678

Institute for Information Systems (IWi)
at the German Research Center for Artificial Intelligence (DFKI)
Saarland University, Campus Building D3 2, 66123 Saarbrücken, Germany
Phone: +49 681 85775 – 3106, Fax: +49 681 85775 – 3696
E-Mail: iwi@iwi.uni-sb.de, URL: http:/ / www.iwi.uni-sb.de/

October 2015

Abstract

In spite of the current research activities developing methods and techniques for business process model analysis, an adequate and homogeneous data basis is still missing. Referring to other research disciplines like computer linguistics or biology, a comprehensive, standardized and digitally available data corpus may lead to a better and consistent understanding of businesses in different domains as well as of the corresponding business information systems. Especially in the context of business process management, such a corpus is of high importance as it improves the creation of particular business process landscapes and the development of standardized evaluations. Against that background, this article presents the IWi Process Model Corpus extending the Reference Model Catalogue, which was developed by the IWi in 2006, with processable process model data. Thereby, its characteristics as well as correspondences between the contained sub corpora and between different single models are focused. In the current version, the corpus contains reference models, models from practice and models from controlled modelling environments and, in total, comprises 24 model collections with 4,426 process models and mappings for more than 137,000 model pairs.

Keywords: process model, process model collection, process model corpus, process model matching

 October 2015

Contents

Abbreviations

AML	ARIS Markup Language
BPM	Business Process Management
BPMN	Business Process Modeling and Notation
DFKI	Deutsches Forschungszentrum für Künstliche Intelligenz (German Research Center for Artificial Intelligence)
e.g.	exempli gratia
EPC	Event-driven Process Chain
EPML	EPC Markup Language
IS	Information Systems
IWi	Institut für Wirtschaftsinformatik *(Institute for Information Systems)*
NSCM	N-Ary Semantic Cluster Matching
PMC	Process Matching Contest
PNML	Petri-Net Markup Language
RMM	RefMod-Miner
XML	Extensible Markup Language

Figures

Tables

1 Introduction

Nowadays, companies use large model databases to manage their business process models, which serve as a knowledge base for the design of their information systems. Oftentimes, these databases contain several hundreds or even thousands of models,[1] wherefore methods and techniques for complexity reduction, handling and analysis of these data are needed. But also for the design of enterprises, comprehensive model collections containing reference models as well as individual models from companies might be helpful in order to derive adequate solutions for particular further companies.

However, access to real process models from practice is missing, which is often caused by legal aspects or privacy concerns. Companies are afraid of losing their competitive advantage by publishing their business processes. In fact, there are several approaches focusing on the conceptualization and the establishment of open access model repositories[2] (apromore.org, openmodels.org, openmodels.at, prozoom.ch) and also initiatives as e.g. Free Models Initiative[3] trying to collect and to spread model data as open data with the corresponding licenses. At the same time, concrete digital and processable models are still very rare[4].

Some trends in that direction can already be observed within the information systems research, e.g. in terms of the interest of the Business Process Management Conference (BPM) in publishing the source code of software tools and implemented algorithms named in the proceedings. In that context, the possibilities of replicating the published findings are of major interest. Nevertheless, publishing the underlying data material is rarely focused. Though, particularly these data are essential for the replication and therefore of high importance for the research progress. The capabilities of corresponding corpora can be observed in different fields of research. E.g. the use of speech and text corpora in the fields of computational linguistics led to high benefits in speech processing, human computer interaction and automatic translation techniques.[5] The use of genomic databases caused substantial progresses in the fields of biology, chemistry and medicine.

Against that background, the paper at hand makes a first step towards a comprehensive process model corpus containing process models in a standardized, digital and processable format. The initial starting point for that aim is the currently existing reference model catalogue (rmk.iwi.uni-sb.de/).[6] It contains 98 reference model entries with lexical data and meta data like the number of contained single models. However, this catalogue does not contain digitally processable models (in terms of the used modeling language or a consistent exchange format) and there are also no

[1] cf. DIJKMAN ET AL. (2011); HOUY ET AL. (2011)
[2] cf. KOCH ET AL. (2006)
[3] cf. THALER ET AL. (2014b)
[4] cf. FRANCE ET AL. (1998)
[5] cf. KUNZE (2005); MILLER (1995)
[6] FETTKE ET AL. (2002)

entries of individual models from different domains. Thus, the authors developed a procedure model which serves as the basis for the extension of the reference model catalogue to the IWi Process Model Corpus. The following research objectives are focused: (1) Creating a consistent understanding of business application systems in different domains, (2) reusing the contained models in other contexts, (3) creating a homogeneous data basis for different application and analysis scenarios. In that context, there is also a wide range of application and analysis methods, for whose (further) development the indented process model corpus is highly beneficial. Some of them are (1) process matching[7], (2) analyzing structural analogies[8] or (3) the search of process variants. The application of existing techniques to a comprehensive model corpus might answer manifold questions, as e.g.: (1a) To which extent are automatic approaches able to find maps which are determined manually? (1b) Are there elements or model fragments which are available in several reference models? (2a) Which structures can be observed frequently, which ones seldom? (2b) Are there different structures in different domains? (2c) Is it possible to define generic process templates? (3a) How does the evolution of models over several years look like?

Finally, the authors aim at publishing the corpus in terms of open models; similar to the open source idea, which was established in the context of software development during the last years. The paper at hand gives an overview on the current version of the developed corpus as well as on additional data material like concrete mappings between the models.

After that introduction, section 2 briefly describes the general procedure model for the creation of the process model corpus. The corpus itself is then described in section 3, where the contained models are characterized with established metrics and additional information in the form of concrete mappings between the single models are presented. Finally, section 4 discusses the results, addresses the release process with continuous integration aspects and closes the paper with an outlook on future work.

2 Procedure Model for Corpus Development

In the context of the paper at hand, a corpus is defined as a structured and versioned library of models and model collections. Model collections, e.g. the SAP-R/3 reference model, cover several single models (in case of the named SAP-R/3 reference model: 604).

In order to develop a model corpus, the authors developed a procedure model (Figure 1) covering its specifications and construction. This includes the model selection, the preprocessing of different sources as well as the process of gathering, harmonizing and providing the models. The procedure model is briefly described in the following. However, details on all phases and its manifestations are presented in WALTER ET AL. (2014).

---

[7] cf. CAYOGLU ET AL. (2013)

[8] cf. EKANAYAKE ET AL. (2012); FETTKE ET AL. (2005); WALTER ET AL. (2012)

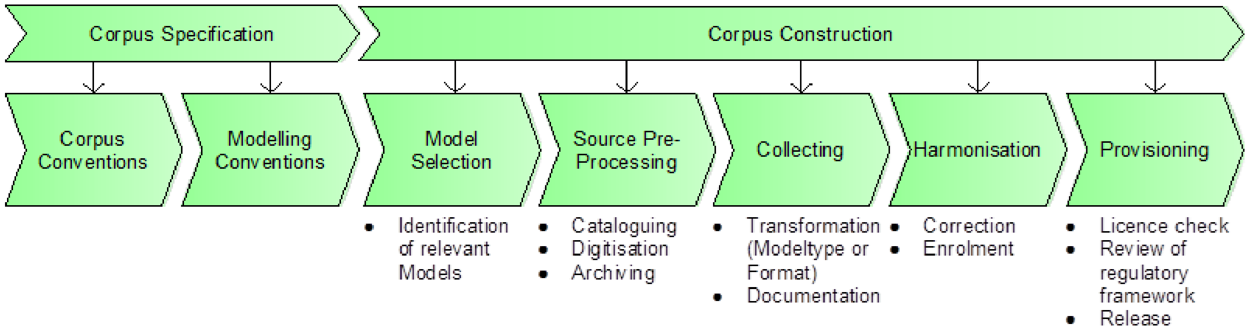

Figure 1: *Procedure model for corpus development*

Corpus Specification. The concrete implementation of the corpus is controlled by the definition of corpus conventions and modeling conventions. The corpus conventions cover all rules and general requirements which affect the whole corpus as e.g. the included model types, the modeling languages and particular transformation procedures between modeling languages as well as the exchange format of the corpus and the version control mechanism. With that background, the IWi Process Model Corpus contains Process Models, generally as EPCs and uses the ARIS XML-Format as exchange format.

Corpus Construction. Based on the corpus specification, the corpus construction contains several steps for the development and the management of the corpus. In the **model selection** step, potential model containing documents of arbitrary formats like text (e.g. books, scientific papers, journals), audio (e.g. interview recordings) or digitized models (respectively process models of different languages and exchange formats) are identified, selected and prioritized for the corpus inclusion. In the context of the **source pre-processing**, the models are cataloged and transformed into a digital and processable format. Beyond that, **collecting** includes the transformation of the models to the defined modeling language and exchange format as well as rework concerning the corpus conventions. In the phase of **harmonization**, the collected models, which are already available in a unified modelling language and in a single target format, have to be transformed with regard to the modelling conventions. This step results in two different variants, “original” and “adapted”, of a sub corpus. If no violations occur, the collected “original” model is equal to the harmonized model. Thus, an adapted model does not exist. For **provisioning**, the legal framework has to be considered before any publication of the models. This typically affects the licensing law as well as the copyright of the sources’ authors.

3 Corpus Overview

3.1 Scope and Classification

The IWi Process Model Corpus contains process models from different application domains. Some of the major domains are e.g. industry, retail and public administration. These models have been collected over many years within BPM research.

Table 1: *Corpus Overview*

Sub Corpus	EPC	BPMN	Remarks	L	C	Source
Becker 2012	7		7 test models	-	I	BECKER ET AL. (2012)
Bizagi		17	17 reference models in the context of system engineering modeled as EPC and BPMN	en	R	http://www.bizagi.com/en/community/process-xchange
Custom B2B	33		Processes describing software customizing and the production of special machinery.	de	R	research project
GK-Rewe	105		Basic course „accounting“ at Chemnitz University	de	I	KAHLERT (2010)
ECO-Integral	85		Digitalization of processes in the context of environmental management	de	R	KRCMAR ET AL. (2000)
E-Payment	38		Processes in the context of the E-Payment project.	de	I	research project
External/anonymized	457		Anonymized process models. Certain conference papers, theses and dissertations	de en	I	anonymized
Business registration	24		Business registration processes of 8 German communes.	de	I	research project
Retail-H-Model 1996	54		Retail information systems. Edition 1996. Contains 54 EPCs and 2 event hierarchies (as EPC).	de	R	BECKER ET AL. (1996)
Retail-H-Model 2004	237		Retail information systems. Edition 2004. Contains 58 EPCs and 2 event hierarchies (as EPC).	de	R	BECKER ET AL. (2004)
ITIL	18		Reference model for the IT Service Management. Contains 19 EPCs, with an example for explanation, and further 297 models of other types.	de en	R	bought from Software AG, OFFICE OF GOVERNMENT COMMERCE (2010a); OFFICE OF GOVERNMENT COMMERCE (2010b); OFFICE OF GOVERNMENT COMMERCE (2010c); OFFICE OF GOVERNMENT COMMERCE (2010d); OFFICE OF GOVERNMENT COMMERCE (2010e)
Exams	78		Exams of a course at a German University between 2010 and 2012.	de	CM	exams
Kurbel	24		Digitalization of processes in the context enterprise resource planning.	de en	I	KURBEL (2011)
Lichtenegger	74		Dissertation of Lichtenegger	de	R	LICHTENEGGER (2012)
Process Matching Contest 2013	90		Birth registration processes of 9 countries and University admission processes of 9 German Universities. Originally modeled as Petri-Nets. PNML files were transformed to EPCs with ProM.	en	I	CAYOGLU ET AL. (2013)
SAP	2,416		SAP R/3 reference model with and without cryptic model names and with and without hierarchies.	en	R	IDS Scheer 1999, ARIS for R/3 Version 4
SAP Custom Dev.	22		Processes in the context of individual software development.	de	I	research project
SAP R/3	56		SAP R/3 reference model. literal, syntactical and referencing errors corrected.	de	R	KELLER ET AL. (1998)
Vogelaar	401		Dutch governance processes. Originally modeled with YAWL. Transformed to EPCs using the transformation rules from the source document.	en	I	VOGELAAR ET AL. (2012)
Y-CIM 1.0	7		Structural correspondent to the German Y-CIM 1998.	en	R	ARIS Toolset 1.0
Y-CIM de	100		Reference model for industrial business processes. Covers EPCs and function trees; inclusive exercise EPCs and descriptions.	de	R	SCHEER (1994)
Y-CIM en	100		Structural correspondent to the German Y-CIM 1998.	en	R	SCHEER (1998)
Number of all models	**4,426**	**17**				

Legend: L: language; C: category (R: reference model, I: individual model, CM: controlled modeling)

The process model corpus is continuously extended by adding further models from both researchers at the IWi and external contributors. The IWi Process Model Corpus contains process models in a standardized, digital and processable format, which are structured into different sub corpora. These sub corpora contain at least one process model. The differentiation into particular sub corpora helps to provide a quick aggregated overview of the different process types and consolidates information in certain metrics (for more information see section 3.2. Table 1 shows some statistics about the whole corpus. The first column depicts a listing of the names of each particular model. The following columns show, besides the model type (EPC, BPMN and Function Tree), the format of the model (AML, EPML, PNML, BPMN), and the classification (individual-model, reference-model and models from controlled modeling scenarios). The classification of the models is based on their origin and type. Thus, each sub corpus within the developed model corpus can be assigned to one of the following three categories[9]:

- Reference Models: Reference models generally consist of descriptive and prescriptive model elements:[10] In a descriptive sense, a reference model captures similarities of a category of companies. In a prescriptive sense, a reference model presents a proposal for the design of enterprises.
- Individual Models: Individual models describe processes in specific organizations. This includes process models from existing companies as well as from public administration.
- Models from controlled modeling scenarios: These models emerge from controlled modeling scenarios, where different test persons are provided with a textual description of a procedure or process. Thus, all test persons can draw on a common understanding of the problem as well as on a uniform terminology. This textual description helps the test persons to have both a common understanding of the problem and a uniform terminology.

Table 1 gives an overview of the current status of the corpus.[11]

3.2 Characteristics of the corpus

The IWI Model Corpus can generally be analyzed by certain metrics. The present paper considers 30 metrics, which are widely used to analyze process models[12]. The calculated metrics are presented in table 2. These metrics are only calculated for process models and not for model types like function trees etc. The metrics calculated on single process models are discarded within the subfolders of the model corpus.

The calculation in table 2 shows the metrics in an aggregated way. The aggregation is calculated as a weighted arithmetic mean. This mean is calculated with the quantity of the concerned process models in the sub corpus. The calculations have been continously developed in the research group at the Institute for Information Systems at the German Research Center for Artificial Intelligence.

[9] cf. WALTER ET AL. (2014)
[10] cf. BROCKE ET AL. (2013)
[11] Version 6 of the IWi Process Model Corpus, released in 2015
[12] cf. MELCHER (2012)

Table 2: *Corpus characteristics*[13]

	number start events	number internal events	number end events	number events	number functions	number AND splits	number AND joins	number XOR splits	number XOR joins	number OR splits	number OR joins	number connectors	number nodes	number edges	density(1)	density(2)	coefficient of connectivity	coefficient of network complexity	cyclomatic number	avg. connector degree	max. connector degree	separability	sequentiality	depth	mismatch	heterogeneity	token splits	control flow complexity	join complexity	weighted coupling
Becker 2012	1.0	0.0	1.0	2.0	9.5	0.0	0.0	1.8	1.8	0.3	0.3	4.0	15.5	18.1	0.1	0.2	1.2	21.3	3.6	3.4	4.1	0.5	0.3	1.0	0.0	0.0	0.5	6.6	7.1	0.1
Custom B2B	4.2	37.9	6.2	46.5	44.7	1.8	2.0	7.8	5.3	0.2	0.2	16.5	107.6	126.5	0.0	0.1	1.1	150.5	21.0	5.5	19.7	0.2	0.3	1.3	5.7	0.4	7.2	564.7	566.6	0.0
GK-Rewe	1.3	13.4	0.5	15.2	10.1	3.2	3.1	1.4	1.3	0.9	0.9	10.7	36.0	42.7	0.0	0.1	1.2	50.8	7.7	3.6	5.2	0.4	0.2	1.7	2.7	0.6	7.4	17.0	18.8	0.0
ECO-Integral	1.1	3.5	1.1	5.7	6.4	0.0	0.1	0.4	0.1	0.4	0.5	1.5	13.6	13.1	0.1	0.0	0.9	12.8	0.7	1.6	1.7	0.7	0.6	0.2	1.3	0.1	0.6	2.6	2.4	0.1
E-Payment	19.8	19.0	20.5	55.3	183.7	2.3	2.0	5.1	5.3	0.0	0.0	16.0	255.0	217.6	0.0	0.1	0.8	187.2	15.0	12.1	72.9	0.0	0.0	0.2	7.5	0.5	58.9	37.2	35.3	0.0
External/anonymized	1.5	8.9	2.0	12.4	8.8	0.9	0.9	2.0	1.2	0.3	0.6	6.0	27.2	29.2	0.1	0.1	1.0	31.6	3.0	3.1	3.6	0.5	0.4	1.0	4.2	0.4	1.9	10.0	10.0	0.1
Business registration	11.7	6.0	12.3	23.0	76.3	1.3	1.3	2.7	2.3	0.0	0.0	9.0	108.3	61.0	0.0	0.0	0.5	37.3	0.0	6.7	30.3	0.0	0.0	0.0	9.0	0.4	19.0	13.3	15.3	0.0
Retail-H-Model 1996	0.7	0.6	0.7	0.4	0.5	1.1	0.9	0.8	0.8	2.0	1.6	0.6	0.5	0.5	0.8	0.5	0.1	0.6	0.8	0.2	0.3	0.4	0.5	0.7	0.6	0.6	1.1	1.2	1.0	0.9
Retail-H-Model 2004	1.4	6.7	1.4	8.7	6.0	1.1	1.0	2.1	2.0	1.2	1.1	5.1	19.3	20.8	0.4	0.3	0.6	22.6	2.5	1.7	2.2	0.4	0.4	1.1	2.3	0.6	1.5	6.4	6.1	0.5
ITIL	6.0	11.0	6.0	23.0	98.0	1.0	1.0	5.0	6.0	0.0	0.0	11.0	132.0	85.0	0.0	0.0	0.6	54.7	0.0	9.1	35.0	0.0	0.0	1.8	11.0	0.4	15.0	28.0	37.0	0.0
Exames	1.3	13.9	2.9	18.1	14.4	0.6	0.5	3.9	1.8	0.3	0.4	7.6	40.1	42.0	0.0	0.1	1.0	44.3	3.6	3.1	3.4	0.4	0.5	1.3	6.5	0.3	1.0	10.1	6.5	0.0
Kurbel	1.6	9.8	1.2	12.6	9.5	1.6	1.6	0.6	0.6	0.1	0.3	4.7	26.8	27.6	0.0	0.1	1.0	28.5	1.8	3.0	3.3	0.7	0.5	1.0	1.0	0.4	1.7	3.0	3.9	0.0
Lichtenegger	1.7	8.2	1.4	11.4	9.0	0.3	0.6	3.3	3.5	0.3	0.2	8.1	28.5	33.5	0.1	0.1	1.1	40.0	6.0	3.2	4.2	0.5	0.4	0.8	3.8	0.3	1.1	13.0	10.8	0.1
Process Matching Contest 2013	1.0	27.8	1.1	29.9	28.8	1.5	1.1	5.5	4.4	0.2	0.4	13.1	71.8	79.8	0.0	0.1	1.1	88.7	8.9	3.3	5.1	0.2	0.5	2.1	5.7	0.4	3.1	25.4	35.5	0.0
SAP	3.9	3.1	4.5	11.5	4.0	1.1	1.1	0.9	1.0	0.6	0.5	5.2	20.7	20.8	0.1	0.0	0.9	21.1	1.4	3.3	4.4	0.6	0.3	0.5	6.0	0.4	3.2	1188.0	179.5	0.1
SAP Custom Dev.	1.0	8.1	1.4	10.5	7.9	0.0	0.0	2.1	1.7	0.0	0.0	3.8	22.2	23.1	0.1	0.1	1.0	24.2	1.9	2.8	3.0	0.5	0.6	0.5	0.8	0.0	0.0	4.4	3.7	0.1
SAP R/3	0.9	14.1	1.8	16.8	12.6	2.4	2.0	2.3	2.0	1.1	1.2	11.0	40.4	46.1	0.0	0.1	1.1	53.0	6.7	3.8	5.4	0.5	0.2	1.4	5.8	0.7	6.8	56.9	49.1	0.0
Vogelaar	1.0	9.0	1.0	11.1	27.1	1.6	1.5	5.7	5.9	0.4	0.4	15.4	53.5	62.2	0.0	0.1	1.1	72.2	9.6	3.2	4.0	0.2	0.3	3.1	0.6	0.5	3.1	15.6	15.9	0.0
Y-CIM 1.0	51.0	143.0	18.0	212.0	148.0	24.0	13.0	26.0	7.0	11.0	42.0	123.0	483.0	545.0	0.0	0.0	1.1	615.0	63.0	3.6	9.0	0.3	0.2	3.0	152.0	1.0	58.0	155.0	585.0	0.0
Y-CIM de	2.0	4.0	1.2	7.1	5.2	0.5	0.3	0.8	0.3	0.2	1.0	3.0	15.3	15.7	0.1	0.0	0.9	16.4	1.5	3.0	3.5	0.5	0.3	0.3	4.2	0.2	1.0	3.4	29.3	0.1
Y-CIM en	2.0	4.0	1.2	7.1	5.2	0.5	0.3	0.8	0.3	0.2	1.0	3.0	15.3	15.7	0.1	0.0	0.9	16.3	1.5	3.0	3.5	0.5	0.3	0.3	4.1	0.2	1.0	3.3	26.7	0.1

[13] The formal definitions of the metrics are presented in MELCHER (2012).

To support the calculations and also to calculate sub results in a detailed manner, we used python scripts for preprocessing the data.

The preprocessing phase ducted with a java program for reference model mining[14]. This program is continuously splits each dataset of a sub corpus into individual model files to prepare the base for the metric calculation of each single sub corpus. Furthermore one sub corpus is published anonymized. Every model name within this sub corpus was anonymized since these models are only for internal use.

All calculated metrics are based on datasets encoded as AML files. The AML exchange format is based on an XML structure. Furthermore, the metrics are calculated on the level of sub corpora. Sub corpora with more than one model set were aggregated with weighted metric results. The weighting is based on the number of models within the respective model sub corpus. The table below shows the metrics calculated on the IWi Process Model Corpus. More detailed calculations are listed in the ZIP file of the corpus. The Metrics within Table 2 are an extension to the already calculated metrics of the process model corpus in Walter et al. (2014)[15].

On the aggregated level of the metrics shown in the table 2, the differences are slight. Metrics which could not be calculated, for example because of a very high computational complexity or deadlocks, were not referred to in the table. This problem mainly occurred in the calculation of the cross-connectivity metric, the process variants and the number of graph components. For example the attempt to calculate the cross connectivity for the Custom B2B models (see table 1) the calculation complexity leads to a time-overflow. One reason for this time-overflow are the high amount of “OR” connectors within the models. This leads to very high variants in the calculated sequences and node calculation combination. With regard to the graph theoretical NP-complete problem the calculation for a high amount of model elements is too complex to be solved in polynomial time.

The paper at hand aims to a slight overview of the model corpus, so the described critical metrics are not presented in table 2. In contrast to the computational complexity, deadlocks can possibly be resolved by further model transformations. This underlines the necessity of defining appropriate transformation rules. Furthermore the cyclicity value covers the ratio of the number of nodes in loops to the number of all nodes in a

---

[14] RefMod-Miner, URL:http://refmod-miner.dfki.de/cms/

[15] The comparison of the metrics within the paper at hand and Walter et al. (2014) shows slightly different results. The differences of the results are caused by the further development (minor bug fixes and an extension of the calculation engine) of the java program which was used for calculating the metrics. Another reason is the continuous development of the process model corpus. The process model corpus is currently within the development phase of the 7th version. All calculations within the paper at hand are based on the 6th release.

model. Considering these values, a trend of individual models to a higher value than the reference models can be observed. Especially the Retail-H reference model with a value of 0.6 is demonstrative for that. This is due to the fact of hierarchization and decomposition of reference models, which, in most cases, is not done in individual models. Therefore, loops could occur without impact on the CYC value.

Against that background, the assumption of a higher cyclicity in individual models would be false. In fact, the metric is not sensible for that aspect. If hierarchized and decomposed models were transformed to a flat EPC, they would contain cycles, too. In addition, the mean values of the metrics for the reference and individual models were calculated. A comparison of these values provides information on the general properties of the considered models. They show, for example, that individual models are, on average, twice as large as reference models. An exception is the mean of the OR connectors, which are used in individual models only half as often as in reference models.

The calculated metrics provide an initial overview on process models within the IWi Process model corpus. They also are first indicators of the measurement and inspect certain similarities between process models.

3.3 Correspondences

In order to quantify the correspondence (similarity) between process models or between particular subcorpora of the IWi Process Model Corpus, it is necessary to identify correspondences between the nodes of that models. This is generally be called Process Model Matching[16], whereby two schemas (models) are taken as input, referred to as the source and the target, and a number of maps between the elements of these two models are produced based on an particular correspondence.[17] In that context, a node mapping is formalized as follows.

Definition 1 (Mapping). Let $G = (N, E)$ be a business process model with

- N is a non-empty set of nodes and
- E is a non-empty set of edges.

For two business process models $G_1 = (N_1, E_1)$ and $G_2 = (N_2, E_2)$ we define:

- a map $m = (n_1, n_2)$ is a tupel of two corresponding nodes, where $n_1 \in N_1$ and $n_2 \in N_2$.
- $M = \{m_1, \dots, m_n\}$ is called a mapping between G_1 and G_2.

Definition 1 only provides a formalization instead of a definition in the strong sense, which is founded in the fact, that formal criteria for a map are not given. Instead, such mappings are established by human matchers or even by matching algorithms, which

[16] cf. THALER ET AL. (2014a)
[17] cf. RAHM ET AL. (2001)

again differ in their matching approach. However, automatically matching large repositories of thousand or more process models needs a lot of computing time if we want to find a good matching standard by applying different algorithms and parameter combinations on each matching. Therefore, prior finding representative matching parameters that are expected to achieve a good matching quality for all corpus models can help to save a lot of computing time and gives additional information on the matching relations in the corpus.

In this section, we present a reference matching standard for all process matching combinations in the IWi-Corpus. First, we describe our reference matching standard and its parametrization. Then, we demonstrate an evaluation of its quality based on a model sample from the IWi process model corpus. Our reference matcher is then applied on all English models in the model corpus in order to deliver good reference mapping between all models in the corpus. In section 4.5, we disclose relations between our matchings in respect of sub corpus-affiliation and matching intensity.

3.3.1 Representative Matching Standard

Finding a representative matching standard requires a distinguished sample from a diverse process model repository. For our paper, we access the IWi process model corpus (Walter et al., 2014) with about 4,600 different process models. Out of it, the models from the process matching contest 2013[18], a sample of 18 process models, are selected, for which the reference mapping was defined in an often quoted empirical evaluation in CAYOGLU ET AL. (2013) and which comprises of the best-assumed n:m mappings between all models. The 18 models consist of 9 models concerning birth registration and 9 models that are related to University admission. The reference mapping was manually defined by rating the similarity of model pairs on a scale of 1 to 7[19]. Since the reference mapping was manually defined, its quality is uncertain.

On the sample, our NSCM n:m graph matching algorithm for English models (http://rmm.dfki.de/) is applied as a good matching procedure because of its outstanding performance in the process matching contest CAYOGLU ET AL. (2013). At first, a semantic error detection is conducted, where faulty modelling is automatically fixed. Then, all node pairs are compared with a semantic dictionary-based similarity measure.

For the evaluation of process matching techniques, there are three established metrics, which are extensively used in the recent work: precision, recall and f-measure. The intention of that metrics is to quantify the proximity of the produced results to those ex-

[18] CAYOGLU ET AL. (2013)

[19] cf. CAYOGLU ET AL. (2013)

pected[20]. Precision is the fraction of found node maps, that is correct in terms of a reference mapping. Recall is the fraction of the correct node maps that are found.[21] The F-measure is the harmonic mean between precision and recall. We classify each map either true-positive (TP), true-negative (TN), false-positive (FP) or false-negative (FN) as visualized in the following figure.[22]

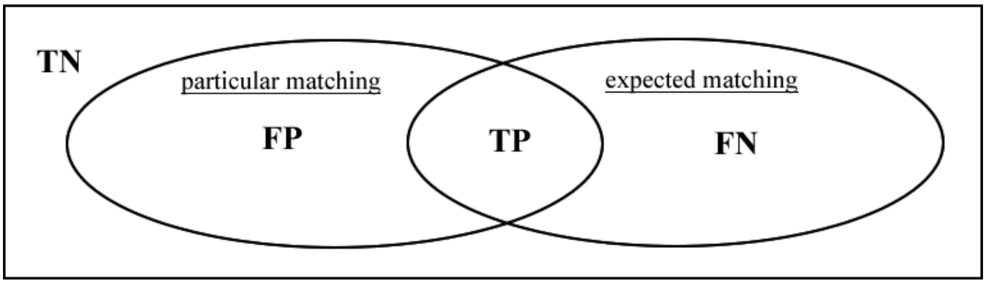

Figure 2: *Relevant sets of mapping for the calculation of precision, recall and F-measure*

Definition 2 (Precision, Recall, F-Measure). Let M_p be a particular mapping and M_r be a reference mapping, such that:

- TP (true positive) is the intersection of M_p and M_r: $TP = M_p \cap M_r$
- TN (true negative) is the complement of M_p union M_r: $TN = \overline{M_p \cup M_r}$
- FP (false positive) is the difference of M_p and M_r: $FP = M_p \backslash M_r$
- FN (false negative) is the difference of M_r and M_p: $FP = M_r \backslash M_p$

Then, $recision = \frac{|TP|}{|TP|+|FP|}$, $recall = \frac{|TP|}{|TP|+|FN|}$ and $fmeasure = \frac{2* precision * recall}{precision + recall}$ is the harmonic mean of precision and recall.

For our quality evaluation, we match all process models in the sample with each other, which results in 153 model combinations. For every matched model combination, its matching quality is recorded by the respective F-measure according to the above mentioned definition of THALER ET AL. (2014a), whereby all mappings are unweighted. Only those mappings were considered that consist of at least one matched node pair. The mappings are available in the *contest-format*.[23] In this file, the first two lines contain the model names and each other line represent a node mapping in the format [node1] | [node2].

[20] cf. EUZENAT ET AL. (2007)

[21] cf. WEIDLICH ET AL. (2010)

[22] cf. CAYOGLU ET AL. (2013)

[23] cf. CAYOGLU ET AL. (2013)

The NSCM n:m matching algorithm performance is outlined in table 2 and 3. The presented values are the arithmetic mean ($mean$) and the standard deviation (SD) of precision ($prec$), recall and F-measure. For our sample, NSCM reached an average precision of 0.8 / 0.37, an average recall of 0.22 / 0.39 and a mean F-measure of 0.31 / 0.34. In comparison with the performance of other matching algorithms in Cayoglu et al. (2013), we believe NSCM to be a good matching algorithm for the application on a huge model corpus.

Table 3: *Precision, recall and F-measure for the PMC models*

dataset	$prec_{mean}$	$prec_{SD}$	$recall_{mean}$	$recall_{SD}$	F_{mean}	F_{SD}
birth registration	0.808	0.194	0.229	0.184	0.319	0.185
University admission	0.373	0.218	0.397	0.274	0.342	0.185

3.3.2 Relations Between Process Mappings

In order to disclose relations between mappings, we match all English models in the IWi Process Model Corpus using NSCM. These models are part of the ITIL, Kurbel, PMC, SAP, Vogelaar and Y-CIM sub corpora. We match only English models because NSCM works only with the English language.

As a result, there are 137,550 model mappings of which 72,514 are empty. The remaining 65,036 mappings consist of 8.99 matched nodes in average with a standard deviation of 9.91. The maximum number of matched nodes is 669. For describing the relations between sub corpora, we define the *matching intensity* as follows:

Definition 3 (Matching Intensity). For a set of business process models $C = \{G_1, \dots, G_i, \dots G_n\}$ (here: the IWi Process Model Corpus) with $G_i = (N_i, E_i)$ and $\forall i,j | i \neq j: N_i \cap N_j = \emptyset$ we define:

- $\Pi = N_1 \cup \dots \cup N_n$ being the union of the pairwise disjoint sets of the therein contained nodes.
- $sc_1 \subset C$ and $sc_2 \subset C$ being sub corpora of C.
- $\mathrm{M}_{sc_1,sc_2} = M_{G_1,G_1} \cup M_{G_1,G_2} \cup \dots \cup M_{G_m,G_n}$ being the union of all mappings between all model pairs between both sub corpora.
- $(n \in \Pi_{sc_1}, m \in \Pi_{sc_2}) \in \mathrm{M}$ with $n \neq m$ so that self-maps are excluded.
- $\Sigma_{SC_1} = \left|\{n \mid (n,m) \in \mathrm{M}_{sc_1,sc_2}\}\right|$ being the number of matched nodes of sub corpus sc_1 to the sub corpus sc_2.

Then, the matching intensity $mi \in [0; 1]$ of two sub corpora is defined as the relative number of matched nodes to the number of all nodes: $mi_{sc_1,sc_2} = \frac{\Sigma_{SC_1} + \Sigma_{SC_2}}{|\Pi_{sc_1}| + |\Pi_{sc_2}|}$.

The resulting mappings are itemized by the matching intensity in Table 4. Figure 3 presents a network that visualizes the matching intensity between the sub corpora in the IWi Process Model Corpus. The node size and redness represents the number of matched nodes of a sub corpus and the edges' thickness represents the matching intensity between source and target node.

Table 4: *Matching intensity between sub corpora in per mille, number matched nodes in brackets*

	ITIL	Kurbel	PMC	SAP	Vogelaar	Y-CIM
ITIL	90 (428)	1 (3)	1 (4)	37 (546)	0 (0)	1 (4)
Kurbel	1 (3)	162 (104)	6 (6)	31 (401)	3 (16)	25 (25)
PMC	1 (4)	6 (6)	318 (821)	45 (625)	11 (63)	31 (62)
SAP	37 (546)	31 (401)	45 (625)	108 (2,710)	30 (510)	47 (622)
Vogelaar	0 (0)	3 (16)	11 (63)	30 (510)	430 (3682)	0 (0)
Y-CIM	1 (4)	25 (25)	31 (62)	47 (622)	0 (0)	260 (358)

As one possible interpretation, the matching intensity can be seen as an indicator for model similarity because the more mutual nodes are matched, the more nodes are similar between process models.

Evaluating the sub corpora matching relations using the matching intensity, the models of the Process Matching Contest 2013 are most closely matched to the SAP and Y-CIM models. The Vogelaar models have the fewest common matchings to all other sub corpora. Among themselves, the Vogelaar models are matched most intensively of all sub corpora. The PMC models and the Vogelaar models have the highest self-matching intensity which can be seen as an indicator for a high contentual homogeneity. The ITIL models, inter-matched, have by far the sparseliest matching intensity of all English sub corpora which speaks for a low contentual homogeneity. For the IWi Process Model Corpus, the implication is that most sub corpora have some often matched nodes in common and the rest is dissimilar. That might be caused by the respective domain's specific language respectively terminology. At this point further research will be necessary to investigate the influence of a domain's terminology on its matching intensity.

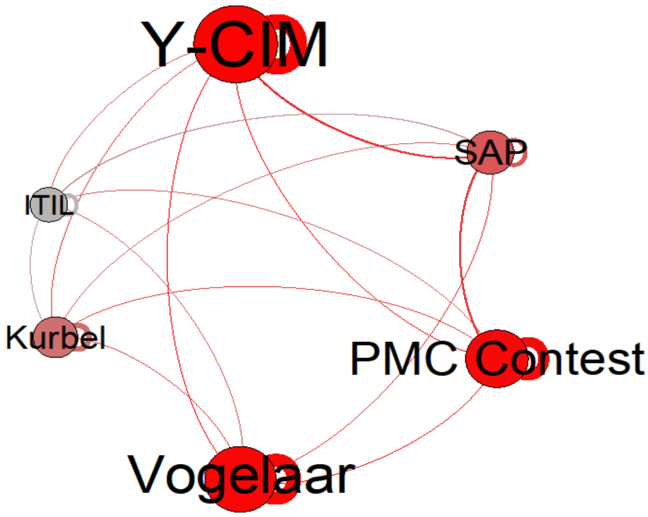

Figure 3: *Network of matching intensity between sub corpora*

4 Conclusion

The presented model corpus was developed based on a procedure model. To achieve a certain basic width, different model collections, such as reference models, individual models as well as models from controlled modelling scenarios, were selected. Moreover, different national languages were considered (Y-CIM, SAP-R/3) and some model collections were taken from various sources (SAP-R/3) as well as from different years of publication (Retail-H, Y-CIM), whereby both analogue and digital sources were considered. Except for ITIL, various changes were made for all reference models in order to be able to provide a consistent data basis. These changes primarily concern the adjustment of syntactic errors identified manually or the transformation of constructs as well as double connectors or sequence operators. Since the original models and the adapted models were added to the corpus with respect to different conventions, different constructs and syntactical rules are used in these models. The simultaneous existence of both model variants allows a wide range for application scenarios.

Although the process corpus contains a high amount of models in different domains, it is narrow in comparison to these domains. Thus, the developed corpus cannot be seen as representative, which can be drawn back to the availability of free accessible models. Against that background it is necessary to continuously extend the process model corpus.

However, the model corpus can be used in a wide range of application scenarios. In order to stress the applicability, three application scenarios are already delivered in the corpus. Within the first scenario, the model corpus was characterized by the use of 30 metrics provided by MELCHER (2012). This calculation is performed in addition to the already calculated results from WALTER ET AL. (2014). These metrics's results are sporadically different. The results discussed in the paper at hand were obtained with the actual version of the java tool RefMod-Miner. The datasets used in the past do not differ from those which are used in the actual calculations. Within the second scenario, all English models were matched to each other using the outstanding matcher (RefMod-Mine – N-Ary Semantic Cluster Matching) of the Process Matching Contest 2013[24]. These mappings were used in the third scenario in order to determine the similarity (in terms of the matching intensity) between all sub corpora. We find the SAP sub corpus to be most similar to the Process-Matching-Contest and Y-CIM models. The Vogelaar and PMC models seem to have a higher contentual homogeneity than the other English sub corpora.

Thus, the authors have taken a first step towards the realization of the presented vision of an extensive model corpus. In contrast to existing approaches, the scientific need for concrete digitally processable models has been addressed, since, in many cases, a lack of a uniform data basis exists. Altogether, the model corpus consists of 24 model collections with 4,426 single models and contains mappings for more than 90,000 model pairs. The contained models cover different domains, characteristics and national languages. Since the corpus is continuously extended and improved, it is published on-line at http://refmod-miner.dfki.de in regular intervals. However, due to legal aspects, it is currently not possible to publish the whole corpus. This limitation especially affects the model data since the models' authors need to agree for a publication as open models. In contrast to that, all additional contents like the detailed metrics, meta-data and the mappings are published as open data within the corpus. In order to further develop the corpus, also external researchers are invited to contact the authors and send feature requests.

The addressed complexity of the developed model corpus enables both the evaluation of existing algorithms, methods and techniques and their (further) development. Here, some possible application scenarios have been outlined briefly, which should be investigated in more detail in future work.

[24] CAYOGLU ET AL. (2013)

References

Becker, J., and Schütte, R. (1996): Handelsinformationssysteme; verlag moderne industrie, Landsberg/Lech; 1996.

Becker, J., and Schütte, R. (2004): Handelsinformationssysteme. Domänenorientierte Einführung in die Wirtschaftsinformatik; Redline Wirtschaft, Frankfurt am Main; 2. Ed.; 2004.

Becker, M., and Laue, R. (2012): A comparative survey of business process similarity measures; Computers in Industry, Vol. 63. 2012, I. 2; pp. 148-167.

Brocke, J.v., and Feetke, P. (2013): Referenzprozessmodellierung;" in: *Enzyklopädie der Wirtschaftsinformatik*, 2013.

Cayoglu, U., Dijkman, R., Dumas, M., Fettke, P., García-Bañuelos, L., Hake, P., Klinkmüller, C., Leopold, H., Ludwig, A., Loos, P., Mendling, J., Oberweis, A., Schoknecht, A., Sheetrit, E., Thaler, T., Ulrich, M., Weber, I., and Weidlich, M. (2013): The Process Model Matching Contest 2013;" in: *4th International Workshop on Process Model Collections: Management and Reuse (PMC-MR'13)*, 2013.

Dijkman, R., Dumas, M., van Dongen, B., Käärik, R., and Mendling, J. (2011): Similarity of business process models: Metrics and evaluation; Information Systems, Vol. 36. 2011, I. 2; pp. 498-516.

Ekanayake, C.C., Dumas, M., García-Bañuelos, L., La Rosa, M., and ter Hofstede, A.H.M. (2012): Approximate Clone Detection in Repositories of Business Process Models; in: *A. Barros, A. Gal and E. Kindler* (Eds.): Business Process Management, 10th International Conference, BPM 2012, LNCS 7481; Springer, Berlin, Heidelberg; 2012, pp. 302 – 318.

Euzenat, J., and Shvaiko, P. (2007): Ontology Matching; Springer-Verlag New York, Inc. 2007.

Fettke, P., and Loos, P. (2002): Der Referenzmodellkatalog als Instrument des Wissensmanagements - Methodik und Anwendung; in: *J. Becker and R. Knackstedt* (Eds.): Wissensmanagement mit Referenzmodellen. Konzepte für die Anwendungssystem- und Organisationsgestaltung; Springer, Berlin 2002, pp. 3-24.

Fettke, P., and Loos, P. (2005): Zur Identifikation von Strukturanalogien in Datenmodellen – Ein Verfahren und seine Anwendung am Beispiel des Y-CIM-Referenzmodells von Scheer; Wirtschaftsinformatik, Vol. 47. 2005, I. 2; pp. 89-100.

France, R.B., Frank, U., Oberweis, A., Rossi, M., and Strecker, S. (1998): Open Models as a Foundation of Future Enterprise Systems (Dagstuhl Seminar 12131); Report Nr., Leibniz-Zentrum für Informatik, Schloss Dagstuhl, pp. 67-85.

Houy, C., Fettke, P., Loos, P., van der Aalst, W., and Krogstie, J. (2011): Business Process Management in the Large; Business & Information Systems Engineering, Vol. 3. 2011, I. 6; pp. 385-388.

Kahlert, D. (2010): Grundkurs Rechnungswesen;" Technische Universität Chemnitz, Chemnitz, 2010.

Keller, G., and Teufel, T. (1998): SAP R/3 prozeßorientiert anwenden – Iteratives Prozeß-Prototyping zur Bildung von Wertschöpfungsketten; Addison-Wesley, Bonn et al.; 1998.

Koch, S., Strecker, S., and Frank, U. (2006): Conceptual Modelling as a New Entry in the Bazaar: The Open Model Approach; in: *E. Damiani, B. Fitzgerald, W. Scacchi, M. Scotto and G. Succi* (Eds.): Open Source Systems, IFIP 203; Springer, Berlin; 2006, pp. 9-20.

Krcmar, H., Dold, G., Fischer, H., Strobel, M., and Seifert, E.K. (Hrsg.) *Informationssysteme für das Umweltmanagement - Das Referenzmodell ECO-Integral*. Oldenbourg, München, Wien, 2000.

Kunze, C. (2005): Semantische Relationstypen in GermaNet; in:Semantik im Lexikon; 2005, pp. 161-178.

Kurbel, K. (2011): Enterprise Resource Planning und Supply Chain Management in der Industrie; Oldenbourg, München; 7. Ed.; 2011.

Lichtenegger, W. (2012): Methoden zur teilautomatischen Konstruktion von lst-Prozessmodellen mittels Process Mining sowie zur Integration manuell konstruierter und automatisch generierter lst-Prozessmodelle; Logos Verlag, Berlin; 2012.

Melcher, J. (2012): Process Measurement in Business Process Management – Theoretical Framework and Analysis of Several Aspects; KIT Scientific Publishing, Karlsruhe; 2012.

Miller, G.A. (1995): WordNet: a lexical database for English; Communications of the ACM, Vol. 38. 1995, I. 11; pp. 39-41.

Office of Government Commerce (2010a): ITIL - Continual Service Improvement; TSO Information & Publishing Solutions, Norwich; 2010a.

Office of Government Commerce (2010b): ITIL - Service Design; TSO Information & Publishing Solutions, Norwich; 2010b.

Office of Government Commerce (2010c): ITIL - Service Operation; TSO Information & Publishing Solutions, Norwich; 2010c.

Office of Government Commerce (2010d): ITIL - Service Strategy; TSO Information & Publishing Solutions, Norwich; 2010d.

Office of Government Commerce (2010e): ITIL - Service Transition; TSO Information & Publishing Solutions, Norwich; 2010e.

Rahm, E., and Bernstein, P.A. (2001): A survey of approaches to automatic schema matching; The Very Large Database Journal, Vol. 10. 2001, I. pp. 334-350.

Scheer, A.-W. (1994): Wirtschaftsinformatik - Referenzmodelle für industrielle Geschäftsprozesse; Springer, Berlin et al.; 4. Ed.; 1994.

Scheer, A.-W. (1998): Business Process Engineering - Reference Models for Industrial Enterprises; Springer, Berlin et al.; 1998.

Thaler, T., Hake, P., Fettke, P., and Loos, P. (2014a): Evaluating the Evaluation of Process Matching Techniques;" in: *Multikonferenz Wirtschaftsinformatik 2014 (MKWI), D. Kundisch, L. Suhl and L. Beckmann* (Eds.), GITO-Verlag, Paderborn, 2014a, pp. 1600-1612.

Thaler, T., Walter, J., Ardalani, P., Fettke, P., and Loos, P. (2014b): The Need for Process Model Corpora;" in: *Free Models Initiative (FMI-14),* Vienna, Austria, 2014b.

Vogelaar, J.J.C.L., Verbeek, H.M.W., Luka, B., and van der Aalst, W.M.P. (2012): Comparing Business Processes to Determine the Feasibility of Configurable Models: A Case Study; in: *F. Daniel, K. Barkaoui and S. Dustdar* (Eds.): Business Process Management Workshops, LNBIP 100; Springer, Berlin; 2012, pp. 50-61.

Walter, J., Fettke, P., and Loos, P. (2012): Zur Identifikation von Strukturanalogien in Prozessmodellen;" in: *Multikonferenz Wirtschaftsinformatik (MKWI), D.C. Mattfeld and S. Robra-Bissantz* (Eds.), GITO, Braunschweig, 2012, pp. 1703-1715.

Walter, J., Thaler, T., Ardalani, P., Fettke, P., and Loos, P. (2014): Development And Usage Of A Process Model Corpus; in: *B. Thalheim, H. Jaakkola, Y. Kiyoki and N. Yoshida* (Eds.): Information Modelling and Knowledge Bases XXVI; IOS Press, Amsterdam; 2014, pp. 437-448.

Die Veröffentlichungen des Instituts für Wirtschaftsinformatik (IWi) im Deutschen Forschungszentrum für Künstliche Intelligenz erscheinen in unregelmäßiger Reihenfolge.

Heft 198: Constantin Houy, Johannes Frank, Tim Niesen, Peter Fettke, Peter Loos: Zur Verwendung von Theorien in der Wirtschaftsinformatik – Eine quantitative Literaturanalyse, Dezember 2014

Heft 197: Peter Fettke, Constantin Houy, Philipp Leupoldt, Peter Loos: Discourse-Orientation in Conceptual Model Quality Research - Foundations, Procedure Model and Applications, January 2014

Heft 196: Constantin Houy, Peter Fettke, Peter Loos: Understanding understandability of conceptual models. What are we actually talking about? − Supplement, Juni 2013

Heft 195: Constantin Houy, Markus Reiter, Peter Fettke, Peter Loos: Prozessorientierter Web-2.0-basierter integrierter Telekommunikationsservice (PROWIT) - Anforderungserhebung, Konzepte, Implementierung und Evaluation, Oktober 2012

Heft 194: Isabelle, Aubertin, Constantin Houy, Peter Fettke, Peter Loos: Stand der Lehrbuchliteratur zum Geschäftsprozessmanagement - Eine quantitative Analyse, Mai 2012

Heft 193: Silke Balzert, Thomas Kleinert, Peter Fettke, Peter Loos: Vorgehensmodelle im Geschäftsprozessmanagement - Operationalisierbarkeit von Methoden zur Prozesserhebung, November 2011

Heft 192: Constantin Houy, Peter Fettke, Peter Loos: Einsatzpotentiale von Enterprise-2.0-Anwendungen - Darstellung des State-of-the-Art auf Basis eines Literaturreviews, November 2010

Heft 191: Peter Fettke, Constantin Houy, Peter Loos: Zur Bedeutung von Gestaltungswissen für die gestaltungsorientierte Wirtschaftsinformatik − Ergänzende Überlegungen und weitere Anwendungsbeispiele, November 2010. Heft 191 wurde auch in englischer Sprache herausgegeben: Peter Fettke, Constantin Houy, Peter Loos: On the Relevance of Design Knowledge for Design-Oriented Business and Information Systems Engineering − Supplemental Considerations and further Application Examples, November 2010

Heft 190: Oliver Thomas, Thorsten Dollmann: Entscheidungsunterstützung auf Basis einer Fuzzy-Regelbasierten Prozessmodellierung: Eine fallbasierte Betrachtung anhand der Kapazitätsplanung, Juni 2008

Heft 189: Oliver Thomas, Katrina Leyking, Florian Dreifus, Michael Fellmann, Peter Loos: Serviceorientierte Architekturen: Gestaltung, Konfiguration und Ausführung von Geschäftsprozessen, Januar 2007

Heft 188: Christine Daun, Thomas Theling, Peter Loos: ERPeL - Blended Learning in der ERP-Lehre, Dezember 2006

Heft 187: Oliver Thomas: Das Referenzmodellverständnis in der Wirtschaftsinformatik: Historie, Literaturanalyse und Begriffsexplikation, Januar 2006

Heft 186: Oliver Thomas, Bettina Kaffai, Peter Loos: Referenzgeschäftsprozesse des Event-Managements, November 2005

Heft 185: Thomas Matheis, Dirk Werth: Konzeption und Potenzial eines kollaborativen Data-Warehouse-Systems, Juni 2005

Heft 184: Oliver Thomas: Das Modellverständnis in der Wirtschaftsinformatik: Historie, Literaturanalyse und Begriffsexplikation, Mai 2005

Heft 183: August-Wilhelm Scheer, Dirk Werth: Geschäftsprozessmanagement und Geschäftsregeln, Februar 2005

Frühere Hefte sind unter http://www.iwi.uni-sb.de/frameset/frameset.php?menu=3&target=/publikation/IWi-Hefte/ verzeichnet.

Unter der wissenschaftlichen Leitung von Professor Dr. Peter Loos sind am Institut für Wirtschaftsinformatik (IWi) im Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) mehr als 60 Mitarbeiter im Bereich der anwendungsnahen Forschung beschäftigt. Seit das Institut vor 30 Jahren durch Prof. Dr. Dr. h.c. mult. August-Wilhelm Scheer gegründet wurde, wird hier in Forschung und Lehre das Informations- und Prozessmanagement in Industrie, Dienstleistung und Verwaltung vorangetrieben. Ein besonderer Anspruch liegt dabei auf dem Technologietransfer von der Wissenschaft in die Praxis.

Die interdisziplinäre Struktur der Mitarbeiter und Forschungsprojekte fördert zusätzlich den Austausch von Spezialwissen aus unterschiedlichen Fachbereichen. Die Zusammenarbeit mit kleinen und mittelständischen Unternehmen (KMU) hat einen bedeutenden Einfluss auf die angewandte Forschungsarbeit - wie auch Projekte im Bildungs- und Wissensmanagement eine wichtige Rolle spielen. So werden in virtuellen Lernwelten traditionelle Lehrformen revolutioniert. Das Institut für Wirtschaftsinformatik berücksichtigt den steigenden Anteil an Dienstleistungen in der Wirtschaft durch die Unterstützung servicespezifischer Geschäftsprozesse mit innovativen Informationstechnologien und fortschrittlichen Organisationskonzepten. Zentrale Themen sind Service Engineering, Referenzmodelle für die öffentliche Verwaltung sowie die Vernetzung von Industrie, Dienstleistung und Verwaltung.

Im neuen Standort im DFKI-Anbau am Campus der Universität des Saarlandes werden neben den Lehrtätigkeiten im Fach Wirtschaftsinformatik die Erforschung zukünftiger Bildungsformen durch neue Technologien wie Internet und Virtual Reality vorangetrieben. Hier führt das Institut Kooperationsprojekte mit nationalen und internationalen Partnern durch: Lernen und Lehren werden neu gestaltet; Medienkompetenz und lebenslanges Lernen werden Realität. Zudem beschäftigen sich die Mitarbeiterinnen und Mitarbeiter mit dem Einsatz moderner Informationstechniken in der Industrie. In Kooperation mit industrieorientierten Lehrstühlen der technischen Fakultäten saarländischer Hochschulen werden Forschungsprojekte durchgeführt. Hauptaufgabengebiete sind die Modellierung und Simulation industrieller Geschäftsprozesse, Workflow- und Groupware-Systeme sowie Konzepte für die virtuelle Fabrik.

Campus D3 2
D-66123 Saarbrücken
Tel.: +49 (0) 681 / 302 - 3106
Fax: +49 (0) 681 / 302 - 3696
iwi@iwi.uni-sb.de
www.iwi.uni-sb.de
www.dfki.de