Preface

MÁRIA RAFFAI
Editor in Chief; Chair of GIKOF/SEFBIS SIG
Professor at Széchenyi University, NR in and Councilor of IFIP,
vice chair of IFIP Information Science TC Enterprise Information Systems WG
eMail: maria.raffai@gmail.hu

The Scientific and Educational Forum for Business Information Systems (SEFBIS) as a Special Interest Group of the John von Neumann Computer Society launched not only a conference series on BIS but publishes scientific results in professional journals both in Hungarian and English languages. The result is 12 Volumes of SEFBIS Journal (in English) and the 11 Volumes of GIKOF Journal (in Hungarian). The present publication is the 13th volume of SEFBIS Journal, it manifests that our efforts have not been in vain, the professionals need a forum for publishing their works. In this year the Editorial Board got altogether 16 papers. The reviewers (blind review from two reviewers) found 10 papers in wide variety of topics from scientific results through business application to professional trainings, that have been worth to publish.

The present volume of SEFBIS Journal contains papers from four different fields of IT:
Research, Scientific Results, Methodologies
- Business processes and knowledge management – a marriage of convenience for the fluid process improvement
- A comparative study of Antminer+ and decision tree classification performances
- Solving exercise generation problems using the improved EGAL metaheuristic algorithm
- Digital archive system for utilization of earthquake-related material

Analysis
- Relationship of organizational culture, leadership and cloud computing
- An information systems design approach suitable to the social context in Japan

ICT Solutions in Business
- The impact of Industry 4.0 on the ERPs
- A new approach a framework for risk assessment maturity
- Efficient energy management system optimization of resources at a furniture company

Education
- Gamification in education: designing and implementing a gamified educational on-line tool

This Journal also reports on the BIS Community's main conferences: the OGIK and CONFENIS that have been organized in the last year and calls the attention to the events of 2019 and 2020 that the readers might be interested in.

As we learned from the conferences, from the forums and discussions the readers are interested in
- the concepts, methods and tools supporting the IS development processes,
- the role and the impact of IS/IT on business and on society,
- the solutions that satisfies customers' requirements and ensures security and privacy, and last but not least
- the Education Space focusing to the field of business information systems.

Performing our goals also in the future we call the researchers, professionals, developers to report on their results, the efficient business and/or educational solutions. Concluding with my sincere greeting to the Readers I wish to obtain new ideas, knowledge about effective ICT innovations in business and research results from all over the world!

The electronic version of the Journal is citied and downloadable from EBSCO Database, and from the SEFBIS HomePage: http://raffa6.wix.com/sefbis#!journals

Maria RAFFAI
Editor in Chief

Business Processes and Knowledge Management
A Marriage of Convenience for the Fluid Process Improvement

MATTEO ARRU
PhD student, Szent István University Corvinus University Budapest, MOL Group
eMail: matteo@arru.it

ABSTRACT

In this article, we are presenting an integrated approach where Business Process Management and Ontology-based Knowledge Management Systems are conjoint benefits from each other in an Enterprise Architecture. Using the PROKEX approach and a business case developed in the logistics of a large Oil company, we will demonstrate an innovative way to support process improvement using a Knowledge Fit measurement that can be obtained using this integrated approach. This measurement can further feedback the models and enrich both the organization's process definition that the knowledge base.

Introduction

Several studies [1] confirm that several IT projects flop and in particular, the promise of Business Process Models (BPMs) of being the backbone of an enterprise architecture frequently fails.

- According to Dumas et al. [2], the common reasons for the failure of BPM programs include:
- A sole focus on BPM methods and tools, not on business goals,
- The belief that BPM is the single source of truth,
- BPM projects that are managed as isolated silos, and
- An overall inability to change.

The third reason is ironic because BPMs are often managed in isolation whereas they should be a holistic management concept that stresses the relevance of looking at the relations between activities, people, and technology to deliver products and services. In the typical scenario, the people working on modelling business process perform a clerical job disjoint from the real way process are implemented, without connecting to another relevant dimension of the enterprise and finally without considering the fluidity of the organization and its processes. A stenotic approach to process implementation and organization where those responsible for building BPM are kept far from the fast pace of the operations is not working, whereas is relevant to find a temperature that can tell us to what extent our model is far from the reality.

A similar critique can be extended to other formalisms such as knowledge bases that even more in information systems are represented through knowledge-based systems (KBS) and to Competence Management Systems. All subsystems that we are considering in this business case. The first general criticality is the isolation of the different formalisms one from the other. The second issue is the functional isolation. This is often due to the low role relevance in the organization of those people that are in charge of developing such systems and therefore tend to work in isolation. The third issue is the distance from the reality of the models due to the simplification of the modelling exercise but also due to the fluidity of the actual world. The final issue is, therefore, the inability of the models to develop with the same pace of the underlying world or to be of any support for the improvement of the actual processes.

To deal with those problems we believe that the concept of the "Knowledge Fit" [3] that we developed demonstrate its validity to support the continuous adaptation of both organization processes and the IT architecture's models (in particular BPM and KBS).

It is not our purpose with this article to provide a comprehensive approach to support the process improvement and the alignment of the IT architecture. This may be obtained only using a holistic approach. We, on the other hand, would like to demonstrate with a real business case how one relevant element in process execution, the performers' competence, can be measured and how this measure can feedback[4] enabling process and IT systems improvement. The concept of the Knowledge Fit has been introduced [5] in the context of the PROKEX framework [6]. The primary goal of this approach is creating an environment where from BPM models, the verbose description of processes and tasks can be extracted for further processing [7].

Research Question:

Can the measure of the Knowledge Fit support reorganization[1] with a knowledge capability perspective? Can this approach feedback the IT architecture improvement too?

Answering this inquiry requires to find a working framework that backings the formalisation of the redesign and, in the meantime, bolster a methodical measurement of the knowledge capability for the framework itself. To build up this, we will show how semantic BPMS conjointly used with the PROKEX framework described in chapter 4, and the STUDIO [8] semantic testing platform can give a sound domain to help the simulation of the organization.

The case study will validate that the approach produces a proper measure that provides a valid indication for process improvement, including the application environment used for that measurement. Harland [9] suggests that in the case of studies, the unexpected should come up, and thus, there is potential to make a useful contribution to knowledge, theory and practice.

This research adopts, therefore, a methodology [10] that is largely used in Business Information Systems, and we will experiment along with the following steps explained more in details in Chapter "The PROKEX framework":

- Set-up of a domain-related knowledge base using ontologies and the STUDIO application.
- Development of Business Process Model (BPM) from corporate documentation of a set of processes in the business case
- Extraction of the knowledge concepts from the BPM
- Matching of the concepts in the ontology with the concepts in the BPM (Ontology Matching [4])
- Calculation of the Knowledge Fit measures
- Evaluation of the process improvement insights from the Knowledge Fit.

Resource Allocation in Competence MS

The application of Competence Management Systems to support Resource allocation and Organization optimisation is particularly relevant. Arias et al. [11] published a state of the art in the research area of Human Resource Allocation in BPM and Process Mining. According to this research, Human resource allocation is an emerging research area that has been generating new proposals applied to real case studies. Most of those studies were published from 2011 to 2016 on scientific Journals and conference proceedings. Most of those papers were validation research and evaluation research using either simulation or case studies.

Arena et al. introduce a Human Resource Optimization (HRO) engine which employs semantically enhanced information and Conditional Random Field (CRFs) probabilistic models with knowledge elicited from workers in an industrial context. The system recommends the right person for the right job in real-time for optimising decisions on how to implement and schedule either repeatedly or non-occurring tasks. [12]

Masum et al. propose an intelligence-based Human Resource Information System with some essential features such as Intelligent Decision Support System for decision making and a Knowledge Discovery in Database for knowledge extraction, and others model using knowledge base and model

[1] In our context a reorganization can involve a change in any of the dimensions: People, Processes and Organization. We must also pay attention to the connection between those three elements of a change: the impact that any individual change has on the other.

base. The model has reasoning capability using experience in solving complex, HR problems, including staffing. [13] Xerox Corporation filed a patent application for a method for role-based auto-selection of employees for training associated with skills required in a project.[14]

Whereas in a traditional organization people are concerns to identify the best tool to perform a specific task, Smirnov et al. [15] highlight that in the Industry 4.0 paradigm also the opposite is relevant because of one of the limitations in the design of applications the unpredictability of availability and nature of human resources abilities. They propose a Platform as a Service to enable applications to identify and provide them with the human resource. The system represents competencies using ontologies and allows flexible discovery of such resources based on availability and knowledge.

The ComProFITS project uses a web-based platform for the evaluation of existing employees and the recruitment of new employees in organizations. [16] This application supports multiple roles, each role can perform several activities, and some activities are provided in more than one role. [17] The application supports the assessment of the employees based on a 360-degree assessment where a team evaluates the competence of the individuals based on the opinion of a group of a co-worker, including subordinates, managers and same-level colleagues. [16]

Bohlouli et al. developed an approach that analyses ComProFit results using statistical analysis of the competences to find the best fitting candidates for specific job positions in companies. Using the Scott-Knott clustering algorithm, it classifies job seekers such as under or over-qualified or best-fit candidates concerning the particular job definition. [18] In this article we are developing a similar approach that is not aiming to identify statistical significance of a specific job fit but rather to provide management with a tool to diagnostic the broader scenario in the absence of the relevant test power. The finding of Bohlouli et al., however, demonstrate the significance of such organizational measurement.

Lili [19] summarizes the most common approaches in the area of human resources optimisation methods. He includes top-down and bottom-up approach[20], 0-1 assignment model [21], multi-project human resource allocation based on the negotiation mechanism with consideration of total cost constraint and individual disciplines [22], M / M / N + M queuing model for call centers [23], " four-in-one " personnel matching method [24], fuzzy input-output optimization model [25], total utility level or cost input condition [26] and proposes an Inverse Optimization Model considering competency disadvantage structure.

The PROKEX framework

The purpose of this chapter is to give an overview of the reference architecture used to implement the PROKEX framework [6]. The framework requires two primary input:

- A BPM representation of the processes existing in the organization
- An Ontology representing the Domain Knowledge

The BPM includes relations between activities but in particular, the roles that should perform the job and a text describing what the job holders should perform. For this exercise, we modelled the processes using Adonis BPM. BPM is generally designed by the process Owners with support of process experts. The other important element is the Ontology. For this exercise, we used STUDIO because the platform embeds a Computer-Based Testing (CBT) system. The Ontology is a semantically enriched graph where each node represents a knowledge element and includes a description.

For the application in the context of this exercise, to each node are associated questionnaires to assess the related knowledge. Knowledge Managers and Domain Experts set-up of the Knowledge Domain ontology.

For the purpose of the calculation of the Knowledge fit we need, in addition to the main components, a description of the Organization model provided by the management and described in the topological tables (see Table 1).

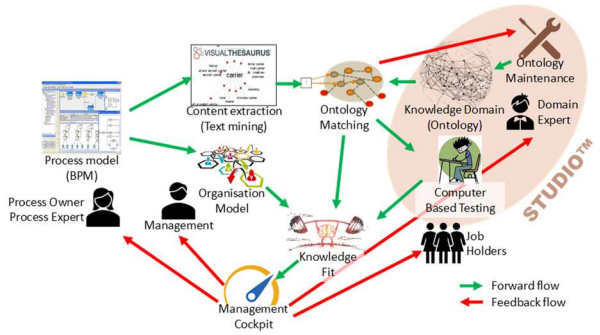

Figure 1. PROKEX's Forward and Feedback flows

The PROKEX framework was initially developed to support the STUDIO CBT and to define a subset of knowledge for test-takers based on the actual job. The process flow to achieve this objective is described in Figure 1 by the green arrows.

In this process, we extract contents by activity using text mining techniques. Those content are mapped using a developed Ontology mapping technique with the Domain Ontology. The result is a concept group that represents the Knowledge required for each role. The result from the tests is valuable to understand not only how the job holders know what they are supposed to know in their position, but by the analysis of the knowledge fit (elaborated in chapter 0) we can support the process of development of policies and processes (Chapter 0) in the organization as concluded in Chapter 0. This feedback flow is represented in Figure 1 using red arrows. If the reader is interested in more details about the research context, I recommend reading the referenced publication [3] and the developed software components [27].

The reference theoretical framework

In this chapter, we summarize the theoretical framework that we are using and therefore, we will referencing to the relevant publications [28] that already discussed the scientific relevance of such an approach in the literature.

The Knowledge Fit measures the gap existing in an organization between the knowledge that we expect to see at a certain level of the organizational deployment [5]. We assume that the knowledge required is fundamentally associated with each activity in the process. It is a standard practice to associate a role with the execution of specific activities; we can, therefore, derive what knowledge is required by each role. Similarly, we associate specific roles to positions that are held by individuals (jobholders/performers). The knowledge required can be defined, therefore at each of those levels. Table 1 shows the topological elements of this organization deployment formalized with binary matrices and vectors.

PROKEX approach introduced the possibility to extract the knowledge elements (using ontology matching [29]) from the description of the processes formalized in BPM. By applying PROKEX, we obtain a vector of knowledge elements that are relevant for the execution of activities and finally, a process. The knowledge required is associated with a specific Concept Group that is a subset of the overall domain ontology [30]. This information is stored in a matrix that can be deployed at all levels of the organization, as illustrated in Table 2.

STUDIO is the component that in the PROKEX framework tests individuals and determines their level of knowledge of a subject included in a specific domain, including those specified in Concept Groups. By testing with STUDIO, the jobholders on the knowledge required as defined by PROKEX, we obtain a Boolean matrix $IT \in \mathbb{B}^{i \times k}$ that contains the information whether the person knows or not the relevant concepts.

Table 3. contains some operators that we used to develop the Knowledge Fit. The Knowledge Fit is initially calculated at the individual level as those that hold a specific knowledge are the jobholders. By association, therefore we can propagate the measurement to the other level of analysis. The Knowledge Fit matrices are analytical measurements of how an element of the organizational deployment (individual, position, role or activity) own or not the knowledge required.

The complementary measure of the Knowledge Fit is the Knowledge Spare. he reader can infer that the knowledge spare shows the knowledge owned that is not required and therefore, an indication of potential flexibility in this point of the organization. To compute the spare knowledge matrices, we can use the formulas in Table 4 replacing the fit function with the spare function.

Table 1. Basic Topological matrices and vectors

Topological Element	Description	Source/Calculation
$\boldsymbol{a} \epsilon \mathbb{T}^{a\times 1}$	List of the positions	From the BPM
$\boldsymbol{r} \epsilon \mathbb{T}^{r\times 1}$	List of the of roles	From the BPM
$\boldsymbol{p} \epsilon \mathbb{T}^{p\times 1}$	List of the positions	From the HR Records of employment
$\boldsymbol{i} \epsilon \mathbb{T}^{i\times 1}$	List of the individuals	From the HR Records of employment
$\boldsymbol{AR} \epsilon \mathbb{B}^{a\times r}$	Activities assigned to Roles	From the BPM
$\boldsymbol{RP} \epsilon \mathbb{B}^{r\times p}$	Roles associated with Positions	From the HR Job description
$\boldsymbol{IP} \epsilon \mathbb{B}^{i\times p}$	Individuals holding a Position	From the HR Records of employment
$\boldsymbol{AP} \epsilon \mathbb{B}^{a\times p}$	Activities associated with a Position	$\boldsymbol{AP} = \boldsymbol{AR} \times \boldsymbol{RP}$
$\boldsymbol{RI} \epsilon \mathbb{B}^{r\times i}$	Roles associated with Individuals	$\boldsymbol{RI} = \boldsymbol{RP} \times \boldsymbol{RI}^{\boldsymbol{T}}$
$\boldsymbol{AI} \epsilon \mathbb{B}^{a\times i}$	Activities associated with Individuals	$\boldsymbol{AI} = \boldsymbol{AP} \times \boldsymbol{IP}^{\boldsymbol{T}}$

Note that we will use the same letter to indicate the primary vector and the related dimension.

Table 2. Matrices of the Knowledge Required

Topological Element	Description	Source/Calculation
$\boldsymbol{k} \epsilon \mathbb{T}^{k\times 1}$	List of the relevant knowledge concepts	PROKEX ontology matching (Concept Group)
$\boldsymbol{AK} \epsilon \mathbb{B}^{a\times k}$	Knowledge required at activities level	PROKEX ontology matching
$\boldsymbol{RK} \epsilon \mathbb{B}^{r\times k}$	Knowledge required at roles level	$\boldsymbol{RK} = (\boldsymbol{AR}^{\boldsymbol{T}} \times \boldsymbol{AK})\ \boldsymbol{AND}\ \boldsymbol{J}_{r,k}$ [1]
$\boldsymbol{PK} \epsilon \mathbb{B}^{p\times k}$	Knowledge required at positions level	$\boldsymbol{PK} = (\boldsymbol{AP}^{\boldsymbol{T}} \times \boldsymbol{AK})\ \boldsymbol{AND}\ \boldsymbol{J}_{p,k}$
$\boldsymbol{IK} \epsilon \mathbb{B}^{i\times k}$	Knowledge required at individual level	$\boldsymbol{IK} = (\boldsymbol{AI}^{\boldsymbol{T}} \times \boldsymbol{AK})\ \boldsymbol{AND}\ \boldsymbol{J}_{i,k}$

[1]Since a typical vector product of two matrices would return not a binary matrix because of the multiplicity in the relationship between activities and roles. For this reason, we will use a "matrix of ones " (also called Unit Matrix [34]) $J_{r,k} = \begin{vmatrix} 1 & 1 & \dots \\ 1 & 1 & \dots \\ \vdots & \vdots & \ddots \end{vmatrix} \epsilon \mathbb{B}^{r\times k}$ to normalise all obtained values to either 1 or 0.

Table 3. Operators needed to elaborate on the Knowledge Fit

Element	Description	Calculation/Notes
$\boldsymbol{A} \epsilon \mathbb{B}^{x\times y}$	Is a generic matrix	$\boldsymbol{a}_{i,j}$ is an element of $\boldsymbol{A}$
$\boldsymbol{v} \epsilon \mathbb{R}^{1\times x}$	Is a generic vector	$\boldsymbol{v}_i$ is an element of $\boldsymbol{v}$
$\boldsymbol{K} \epsilon \mathbb{B}^{x\times k}$	Is a generic matrix of the knowledge required	$\boldsymbol{k}_{i,j}$ is an element of $\boldsymbol{K}$
$\boldsymbol{T} \epsilon \mathbb{R}^{x\times k}$	Is a generic matrix of the knowledge required	$\boldsymbol{t}_{i,j}$ is an element of $\boldsymbol{T}$
$\boldsymbol{fit} \epsilon \mathbb{R}^{x\times k}$	Shows where there are knowledge gaps	$\boldsymbol{fit}_{i,j} = \begin{cases} \boldsymbol{1} & ;\ \boldsymbol{k}_{i,j} = \boldsymbol{0} \\ \boldsymbol{t}_{i,j} & ;\ \boldsymbol{k}_{i,j} = \boldsymbol{1} \end{cases}, \forall \boldsymbol{i} \in [\boldsymbol{1}, \boldsymbol{x}], \forall \boldsymbol{j} \in [\boldsymbol{1}, \boldsymbol{k}]$
$\boldsymbol{spare} \epsilon \mathbb{R}^{x\times k}$	Shows where there is knowledge not requested	$\boldsymbol{spare}_{i,j} = \begin{cases} \boldsymbol{0} & ;\ \boldsymbol{k}_{i,j} = \boldsymbol{1} \\ \boldsymbol{t}_{i,j} & ;\ \boldsymbol{k}_{i,j} = \boldsymbol{0} \end{cases}, \forall \boldsymbol{i} \in [\boldsymbol{1}, \boldsymbol{x}], \forall \boldsymbol{j} \in [\boldsymbol{1}, \boldsymbol{k}]$
$\boldsymbol{elementsByColumn} \epsilon \mathbb{R}^{1\times x}$	Is a vector containing the sum of the values in the columns of a generic matrix	$\boldsymbol{elementsByColumn}_i = \sum_{k=1}^{y} \boldsymbol{a}_{k,i}, \forall \boldsymbol{i} \in [\boldsymbol{1}, \boldsymbol{x}]$
$\boldsymbol{DivideBySize} \epsilon \mathbb{R}^{x\times y}$	Is a matrix where all elements of a column are divided by the value of a specific vector of the same size	$\boldsymbol{DivideBySize}_{i,j} = \frac{\boldsymbol{a}_{i,j}}{\boldsymbol{v}_i}, \forall \boldsymbol{i} \in [\boldsymbol{1}, \boldsymbol{x}], \forall \boldsymbol{j} \in [\boldsymbol{1}, \boldsymbol{y}]$

Table 4. The Knowledge Fit matrices

Knowledge Fit measure	Level of analysis	Calculation/Notes
$\boldsymbol{IFit} \in \mathbb{B}^{i \times k}$	Individuals	$\boldsymbol{IFit} = \boldsymbol{IT\ or\ !\ IK}$
$\boldsymbol{PFit} \in \mathbb{R}^{p \times k}$.	Positions	$\boldsymbol{PFit} = \boldsymbol{fit(PK, dividePerSize(IP^T \times IT, elementsByColumn(IP)))}$
$\boldsymbol{RFit} \in \mathbb{R}^{r \times k}$	Roles	$\boldsymbol{RFit} = \boldsymbol{fit(RK, dividePerSize(RI \times IT, elementsByColumn(RI^T)))}$
$\boldsymbol{AFit} \in \mathbb{R}^{a \times k}$	Activities	$\boldsymbol{AFit} = \boldsymbol{fit(AK, dividePerSize(AI \times IT, elementsByColumn(AI^T)))}$

Table 5. Aggregated Knowledge Fit Scores

Element	Description	Calculation/Notes
$\boldsymbol{F} \in \mathbb{R}^{x \times k}$	Any Fit Matrix	From Table 2.
$\boldsymbol{K} \in \mathbb{B}^{x \times k}$	Any Knowledge required matrix	From Table 4
$\boldsymbol{score_i} \in \mathbb{R}^{x \times i}$.	Knowledge Fit Score	$\boldsymbol{score_i} = \frac{\sum_{j=1}^{k} (\boldsymbol{F} \circ \boldsymbol{K})_{i,j}}{\sum_{j=1}^{k} \boldsymbol{K}_{i,j}}, \forall \boldsymbol{i} \in [\boldsymbol{1}, \boldsymbol{x}]$
$\boldsymbol{gScore} \in \mathbb{R}^{i \times 1}$	Global Knowledge Fit Score	$\boldsymbol{gScore} = \frac{\sum_{j=1}^{x} \boldsymbol{score_j^2}}{\boldsymbol{x}}$

Finally, in Table 5. we introduce the Knowledge Fit Score that provides an average score of the fit for each either individual, position, role or activity. The Global Knowledge Fit Score summaries the level of fit in the particular level with a unique indicator the tells how far the organization is from being optimal. Also, in this case, we can evaluate similarly to the Knowledge Fit Score a Knowledge Spare Score and their related Global Knowledge Spare Score. In Chapter 0, we will see that the "Knowledge Fit" is a tool that is very difficult to read and interpret by a human, whereas Aggregated Knowledge Fit Scores can present relevant indicators for analysis

Business Process Improvement

In the literature, Forster [31] describes "Business Process Improvement" as a systematic approach to help organizations to archive significant changes in the way they do business. The mission of Business Process Improvement methodologies is to focus the process on the creation of value for the customer and to eliminate all that is creating costs without adding value. It is essential to clarify that the criteria for the actual feasibility of a process change cannot only rely on the information coming from the "Knowledge Fit".

A professional analyst when redesigns a process must take into consideration several other aspects including the possibility of implementing the pull, the correct position of the process pacemaker, the attitudinal mix of the resources, the availability of technology, the knowledge required.

In this research, we consider that all the other dimensions have been already addressed and only the resource "Knowledge Fit" need to be analyzed. Otherwise, it is possible to use the "Knowledge Fit" to evaluate a scenario designed using the other criteria with the knowledge perspective.

The measure of the "Knowledge Fit" aims to measure the effect that can show root cause in all the following contexts: The people, the organization, the formalized knowledge, the processes, the corporate knowledge and the measurement system itself [3]. Please note that the development of corporate knowledge and supporting knowledge is a relevant process improvement action [32].

Empirical evidence

Since 2016 it happens to work for the Logistics of MOL Group as Operational Excellence Senior Expert. The MOL Group is a multinational oil company based in Budapest with downstream operations in 8 different countries with 12 companies. Our business case focuses on the logistics of the Italian subsidiary IES.

The business case resulted in being relevant because there is extensive documentation from various company standard, exists a knowledge base, and because IES was passing through a period of continuous reorganization. The expect-

ation from the experiment is to generate a scenario based on the "Knowledge Fit" that confirms or exceed the perception of the local management regarding the following aspects:

- validate the feasibility of the overall assessment process,
- identify those elements that the model is not able to incorporate and
- identify operational issues encountered during the implementation that may open new research streams.

For this business case, we tested the knowledge of 11 employees using the STUDIO platform on 125 concepts extracted from their job role related BPM using PROKEX. Then we elaborate on the "Knowledge Fit" [33] and the "Knowledge Fit Score" as described in the previous chapter..

Hereafter we are discussing the results of the "Knowledge Fit" elaboration from a business point of view rather than in a formal or IT perspective with the expectation that the measurement can tell a story that is relevant from a management point of view. We will show that the results of the analysis will raise attention around people, organization, processes and technology, corporate knowledge or the measurement system as anticipated in Chapter 0. For this analysis, we will mainly use the functions Knowledge Fit Score and Knowledge Spare Score that we defined in Table 6, whereas the detailed elaboration is available in the literature [33].

In the first column, Individual Fit Score Table 6. represents the name of the people[2] who took the test. The following column indicates the number of concepts (of the overall 125 concepts in the Concept Group) that he/she is supposed to know as from the table IK Table 2 followed by the Knowledge Fit Score and Knowledge Spare Scores. As a next step, we will discuss evidence that is emerging, considering the managerial experience that supports or are supported by the analytical outcomes.

Table 6. Individual Knowledge Fit and Spare Score

Individuals	#	Fit Score	Spare Score
Mainteinance2	33	0.67	0.40
Dispatcher2	57	0.42	0.49
Terminal3	27	0.78	0.66
Transportation2	62	0.45	0.62
Transportation1	62	0.65	0.79
Terminal1	27	0.78	0.59
Terminal2	27	0.81	0.44
Retail1	1	0.00	0.50
Wholesale1	3	0.33	0.69
Dispatcher1	57	0.44	0.47
Mainteinance1	33	0.64	0.42

Mainteinance2 is a Maintenance Manager working for IES for more than ten years. He is a mechanical engineer coming from a position in refinery before this was closed. From the test, it results to be proper fitting the position is holding whereas does not score very well in other domains. This is connected to the fact that the domain relevant for maintenance are specific and not so much related to the same logic of distribution and logistics that are in the scope of most of the other activities.

Dispatcher2 is the youngest among the test takers as he finished high school only two years ago. He joined IES recruited from the Race Track to work as a dispatcher for the Racing and Agriculture fuel in one of IES subsidiaries for three months. He is very junior in the position, and that suggests from the result of the test showing that his orientation is not yet completed. He, however, seems to be a bit stronger with technical related jobs, as shown in Table 7. His flexibility is still limited, and this is concluded from a relatively low Knowledge Spare Score. The following $IPFitTab \in \mathbb{R}^{11\times11}$ contains all the Knowledge Fit Score values to all the associations within the IP Table 1. every cell explicit the potential Fit that an individual has with any role (including those that she or he usually not performs).

[2] To maintain privacy, we anonymize the name of the job-holders using aliases.

Table 7. Table of "Knowledge Fit" between individuals and positions

IPFitTab	Dispatcher	Maintenance Manager	Retail Manager	Terminal Manager	Transportation Manager	Wholesale Manager
Mainteinance2	0.37	0.67	0.00	0.63	0.37	0.00
Dispatcher2	0.42	0.64	0.00	0.59	0.44	0.00
Terminal3	0.63	0.76	1.00	0.78	0.66	0.67
Transportation2	0.44	0.79	0.00	0.81	0.45	0.00
Transportation1	0.65	0.73	1.00	0.67	0.65	0.67
Terminal1	0.51	0.79	1.00	0.78	0.55	0.33
Terminal2	0.42	0.79	0.00	0.81	0.44	0.00
Retail1	0.35	0.76	0.00	0.74	0.37	0.00
Wholesale1	0.56	0.76	1.00	0.70	0.56	0.33
Dispatcher1	0.44	0.64	0.00	0.59	0.45	0.00
Mainteinance1	0.44	0.64	0.00	0.63	0.45	0.00
Average	0.48	0.72	0.36	0.70	0.49	0.18

Terminal3 is not working in IES but in the MOL Head Quarter. He was tested because potentially can be a terminal manager although his specialisation is Operational Excellence. He is a computer engineer with an MBA and is working in the industry for two years. However, he had previous experience as Logistics Manager among other positions he holds as specialist and manager. He has one of the highest fit that he makes him a potentially a suitable candidate to hold that position. He has, on the other hand, a high spare score too, and this means that he is flexible to work in different positions. From Table 8 we can assume that the Terminal Manager position has a high Knowledge Fit Score and a relatively high Knowledge Spare Score, too. Therefore, other candidates can hold this position, and *Terminal3* can eventually hold a position where competences are less available.

Transportation2 is a new Transportation Manager Assistant that joined IES one year ago. He has a degree in business and a brief two years' experience in the marketing department of a large multinational consumer goods company. Also, in this case, the indicator shows a lack of business experience. In his case, however, the Spare Score reflects a better fit in other positions those that anyways have several candidates holding the required knowledge. He, in fact, scores best as a Terminal Manager. As he is now developing as a transportation expert that is a technical position, this indication clearly shows that he can acquire the right competence for moving to a Terminal Deputy Manager position in the future.

Transportation1 is with IES since he graduated from a business school 15 years before. He has been the Secondary Distribution Transportation Manager for the last five years. He is the one with the best overall score. However, in the position, he has an average fit. Looking at the following analysis at positions level, we see that in general, the results in this area are weaker than others. *Transportation1* has a high Knowledge Spare Score that is reflecting the long experience he has in the company. According to Table 8, he is one of the few having, for instance, competence in retail and wholesale management.

Terminal1 is a new Terminal Manager in one of IES's subsidiaries. He did not go to university but has almost 20 years' experience in logistics, even if it is the first time working in the oil business. *Terminal1* has a high Knowledge Fit Score and a relatively high Knowledge Spare Score, demonstrating he is a person with the excellent flexibility that is an essential attribute for a position of responsibility such as the Terminal Manager.

Similar conclusions can be drawn for *Terminal2*. He has been Terminal Manager in IES for four years when they closed the refinery and left his position as shift leader of one of the refinery sections. Overall, he was working in the oil and gas industry for 15 years since he graduated from chemical engineering. *Terminal2* scores best in Knowledge Fit even if

his flexibility as reported by the Knowledge Spare Score is relatively low, and this can be explained because his background is more technical than commercial. According to Table 8 he also has the highest fit as Maintenance Manager.

Retail1 is Retail Manager in IES, and he is now managing the sale of the IES Retail network. He will, therefore, soon be reallocated in a different position. He graduated in business and was working as Filling Station Manager for five years before entering in this position three years ago. To understand the absolute low fit score, we will discuss further the Retail Manager Position and its roles. Please note that there is only one concept required for the position as an outcome of the PROKEX iteration. For what concerns the Knowledge Spare Score that is also low, must be said that the type of competencies needed in sales is very different from those that apply to logistics that is mainly represented by this experiment.

Wholesale1 is IES Wholesale Manager for ten years. He held a degree in business and was working in retail until four years ago. He has a low fit score too, and a similar conclusion that we described for *Retail1* can apply to *Wholesale1* that on the other hand, he has a better Knowledge Spare Score. This means that there are different positions where he can better fit. The fundamental reason for this difference that retail works primarily with business-to-customer problems at the filling stations, whereas wholesale is much more integrated with the terminal operations.

Dispatcher1 is the dispatcher for one of the subsidiaries that work mainly with the agriculture business. He works for IES since she left high school 15 years ago and she is Deputy Terminal Manager, too. He has a low level of both Knowledge Fit and Spare Scores this may be a person that requires formal training.

Mainteinance1 is a chemical engineer working for IES for the last 30 years and now is the Chief Maintenance Manager. He is a very energetic person, but very busy. We appreciate that he took the time for taking the test. His result is very similar to *Terminal2's*. When analyzing the Knowledge Fit and Spare Scores about the position, we identify three categories of positions. In the first category, we have Maintenance Managers and Terminal Managers. The Knowledge Fit Score is high for both, and they are both associated with almost 30 concepts. Looking at the IP table, we notice that they have at least two tests each. Further, from table RP, we see that the two positions for two-third they share the same roles.

Table 8. Knowledge Fit and Spare Scores by Positions

Positions	#	Fit Score	Spare Score
Dispatcher	57	0.43	0.48
Maintenance Manager	33	0.65	0.41
Retail Manager	1	-	0.50
Terminal Manager	27	0.79	0.56
Transportation Manager	62	0.55	0.71
Wholesale Manager	3	0.33	0.69

In general, we can conclude that those two positions are adequately staffed, however we can conclude that most of the test-takers were scoring similarly high despite those are the roles usually held by the most experienced people. There are two main reasons for that: the BPM is not detailed enough, or the description is too generic that does not capture the complexity of the role. In this case, it is required to improve the description of the processes. The other reason is that those are managerial jobs where more than the knowledge is needed the capacity to make decisions based on experience and organizational influence. Those are the factors that model like the one that this study is covering focusing only on the knowledge is not able to capture. We see, on the other hand, that the Terminal Managers are resources that are very flexible and there is an opportunity for incorporating different roles that are suffering from missing competence and eventually delegate some activities to the Maintenance Managers. This may happen with a simple reallocation of roles or may require a redefinition of the processes.

The second group of positions are Dispatchers and Transportation Managers. According to the job description, the Transportation Manager should be able to perform the Dispatcher role as well. Hierarchically the Dispatcher is under the supervision of the Transportation Managers. The Knowledge Fit Score resulting from the test is not particularly high. The number of concepts required for those positions is quite high and around 60.

Analyzing the concepts that are related to these positions, we identify a few issues with the testing process and the business process. For instance, between Dispatcher and Transportation Manager, only one managed to reply correctly to the question related to *barges*. According to the STUDIO testing modalities [8], the following questions around barges were not asked showing potential for improving the testing process. Further is not clear why the local IES procedures include a definition of either *rail* and *barge*: barge operations are related only to one of the locations where IES operates, but it is out of scope in this exercise, and no rail operations are happening at all. Requiring the knowledge of that concept is therefore not correct, suggesting that the documentation should be improved, and BPM should be more specific to the activities that are performed in each different site. The results also recommend that those individuals holding those positions must increase particular competence.

In the last group of positions, we identify Wholesale and Retail Managers. The Knowledge Fit Score, in this case, is unusually low. As a premise, we must clarify that the processes in the scope of this exercise are mainly logistics processes. The roles that are relevant to those positions are typically related to sales processes and only residually pertaining to the logistics. It is not optioned, therefore evaluate the fit for those positions and the individuals holding them. What is relevant, instead is to assess the Knowledge Spare Score that shows that *Wholesale1* has the better possibility of being reallocated than *Retail1*. Further, the analysis of the score can give some indication about the corporate knowledge: the knowledge required for those position consists of very few concepts requiring the further elaboration of the description of the business process.

In the analysis of the Knowledge Fit and Spares Scores in Table 10, we would like to clarify that the meaning of the Knowledge Spare Score represents the possibility of enriching the role with new activities among the ones under evaluation. In this context, there is a particular possibility with the business operation that, on the other hand, is connected to non-required knowledge. We saw already that the Terminal Managers and Maintenance Manager hold considerable flexibility but from this analysis results the possibility to use it more in some other roles like, for instance, local sales.

Table 9. Roles Fit and Spare Score

Roles	#	Fit Sc.	Spare Score
Business operation	0	-	0.61
Dispatcher / shift supervisor	47	0.48	0.58
Persons responsible for transportation	5	0.60	0.63
Local Retail	1	-	0.50
Local Sales	3	0.33	0.69
Local product storage mgmnt	3	0.47	0.56
Maintenance management	11	0.77	0.45
Operator	4	1.00	0.54
Quality Control	12	0.65	0.55
SSC	2	-	-
Scheduler	10	0.53	0.54
Technical execution	3	0.73	0.55
Terminal manager	5	0.87	0.60

In this analysis, without repeating what already discussed in the previous chapters, we would like to focus on a few observations.

The low Knowledge Fit Score and good Knowledge Spare Score for the dispatcher role is relevant. IES is planning two potential actions: on one side, there are advanced discussions of outsourcing the truck fleet on which, however, the MOL Head Quarter is not very convinced. On the other hand, there is the possibility of introducing a scheduling tool that may strongly simplify the role of the dispatcher. In both cases, the "Knowledge Fit" supports the re-engineering of the process with technological development.

The second more evident conclusion is related to roles that have limited or null concepts connected. This will be more evident when we will see the fit at the activity level whereas at the role level, on the other hand, is visible the presence of an SSC role (Single Service Company) that has both zero Knowledge Fit and Spare Scores. The reason is that this role is outsourced and therefore not associated with any position.

Table 10. Knowledge Fit and Spare Scores by Activities

Activities	#	Fit Sc.	Spare Score
100. Metrology	3	0.47	0.56
110.1.a Administrative return goods	0	-	0.61
110.1.b Physical Return goods handling	1	1.00	0.61
110.2. Off-spec product management	1	0.67	0.61
120. Reporting	1	1.00	0.61
Acceptance/transfer in tank	0	-	0.54
Calculating necessary transportation capacity	1	1.00	0.63
Checking loading conditions - Train	3	0.42	0.55
Checking loading conditions - Barge	1	0.50	0.54
Checking loading conditions - Road	2	0.63	0.54
Controlling the quality	4	0.63	0.54
Controlling the quantity - Barge	4	0.31	0.55
Controlling the quantity - Pipeline	1	1.00	0.54
Controlling the quantity - RTC	4	0.50	0.54
Controlling the quantity - RoTC	1	-	0.55
Create transportation plan for next day/shift	3	0.75	0.54
Defining optimal replenishment inventory level	3	0.33	0.64
Ensure available capacity	0	-	0.63
Execution of “autonomous maintenance	0	-	0.61
Execution of “routine maintenance	3	0.73	0.55
Inspecting the vehicle	0	-	0.54
Inventory checking (FS & VMI accounts)	1	1.00	0.54
Inventory checking - managing data quality	0	-	0.54
Issuing the transport documents - Barge	1	-	0.55
Issuing the transport documents - Train	2	0.88	0.54
Issuing the transport documents- Road	1	1.00	0.54
Loading the transport means - Train	4	1.00	0.54
Loading the transport means - Barge	0	-	0.56
Loading the transport means - Road	0	-	0.56
Making preparations for product reception, reviewing the transport documents	2	0.83	0.61
Monitor FS (VMI) turnover	0	-	0.63
Notification	0	-	0.48
Order generation for filling stations (VMI accounts)	4	0.25	0.55
Plan-Fact evaluation, controlling	1	1.00	0.63
Planning and Scheduling	2	-	-
Quality control - Barge	0	-	0.56
Quality control - Train	9	0.53	0.56
Quality control- Road	3	1.00	0.55
Receiving Retail business plans, forecasted delivery volumes	0	-	0.50
Receiving Wholesale business plans, forecasted delivery volumes	1	-	0.69
Receiving actual information about delivery fulfilment from hauliers	0	-	0.54
Receiving customer orders, forwarding to R&S	2	0.50	0.68
Receiving daily inventory and sales reports from FSs (and other VMI accounts)	1	-	0.50
Receiving the transport means - Barge	0	-	0.54
Receiving the transport means - Train	6	0.71	0.53
Receiving the transport means- Road	3	1.00	0.53
Record keeping and registration in inventory	1	-	0.55
Sales forecasting	2	0.50	0.54
Sending confirmation of a scheduled delivery time to the local Sales organization and/or customers and/or FS	0	-	0.54
Sending information about transportation plans to haulers and/or terminals	0	-	0.54
Transferring the risk of the product, registering the discharge - Train	1 1	0.23	0.57
Transferring the risk of the product, registering the discharge - Barge	2	0.13	0.55
Transferring the risk of the product, registering the discharge- Road	0	-	0.54
Unloading	0	-	0.56
Verification	5	0.70	0.47
Work order selection (RBWS)	6	0.83	0.46

From Table 10 we can see that several activities do not have associated any ontology concept that suggests that BPM must be further improved. On the other hand, a revision of the ontology may also be necessary to be sure that the ontology matching process identifies the concepts described. It is not the objective of this article discussing the limitations related to the current ontology matching algorithm that could have impacted the weak association of concepts to specific task description in the BPM. This is an indication that an area where we need to

improve the measurement system further. The fact is particularly interesting that for the activity *"Transferring the risk of the product, registering the discharge"* is repeated for the different transportation means, but the matching is opposite to the actual need in the organization. As previously discussed, IES has no rail operations, but the activity related to rail operations matches 11 concepts, have limited barge operations and matches two concepts and continuous road operations whereas matching no concept at all. This is a severe misrepresentation of the formalized business process in comparison with the actual processes.

Results / Recommendations

In response to the research question we propose an approach to process improvement based on Knowledge Fit. We demonstrated how this approach in practice provided IES' management with a series of relevant evidence to support their reorganization. The analysis evidenced the area where both BPM and ontology centric knowledge-based system needed to be further developed or aligned.

The "Knowledge Fit" concept that we experimented in this exercise was developed using PROKEX technology and introduces a systematic translation between the process and knowledge domain and BPM in an integrated approach. This is a novelty compared to different approaches that target is the analysis of knowledge assets to support organizational changes; this includes the increased level of granularity and an integrated knowledge management approach. Further, the traditional approach bases their capability assumption only on the expertise of individuals, whereas with the PROKEX approach those assumptions are supported by a process for elicitation based on the analysis of the process definition artefacts. Further, the evaluation of the current capability is based on actual testing of the staff knowledge. This has the advantage of the objectivity, whereas it is missing to analyze another aspect relevant to evaluate the jobholders' skills (including experience, attitude). The Knowledge Fit further provides management with the possibility of discussing this capability using a quantitative evaluation further than an only subjective evaluation. This is a compelling element because it facilitates the simulation of different organizational scenarios using BPM and to evaluate the impact by a knowledge point of view too.

A relevant area of further development is to use the Knowledge Fit measure to support organization automatization using operational research techniques. The Knowledge Fit scores can be seen as a measure that can be maximized by an optimization engine opening a vast area of applications. The Knowledge fit can inform management decisions together with other measurements (including cultural fit and market fit) to support structured organizations to follow the fluid market without missing the inside constraints.

References

[1] M. Raffai, "International Program for Standardizing Global IT Profession," *SEFBIS J.*, vol. 3, pp. 1–13, 2008.

[2] M. Dumas, M. La Rosa, J. Mendling, and H. A. Reijers, "BPM as an Enterprise Capability," in *Fundamentals of Business Process Management*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2018, pp. 475–500.

[3] M. Arru, "Use of Ontologies and Business Process Management Systems to Measure the 'Knowledge Fit' of an Organization," Budapesti Corvinus Egyetem, 2019.

[4] A. Gábor *et al.*, *Corporate Knowledge Discovery and Organizational Learning*, vol. 2. ham: Springer International Publishing, 2016.

[5] M. Arru, "Developing a Measure of Intellectual Capital fit : an Approach to Improve Business Processes," in *European Conference on Intellectual Capital*, 2016, pp. 382–393.

[6] "ProKex," *Semantics and Corporate Knowledge Management*, 2013. [Online]. Available: http://prokex.netpositive.hu/. [Accessed: 23-Mar-2017].

[7] M. Arru, "Application of process ontology to improve the funding allocation process at the European Institute of Innovation and Technology," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8650 LNCS, pp. 133–147, 2014.

[8] C. Weber and R. Vas, "Studio: Ontology-Based Educational Self-Assessment.," in *Workshops Proceedings of EDM 2015 8th International Conference on Educational Data Mining*, 2015, vol. 1446, pp. 33–40.

[9] T. Harland, "Learning about case study methodology to research higher education," *High. Educ. Res. Dev.*, vol. 33, no. 6, pp. 1113–1122, 2014.
[10] M. Török, "Organizational knowledge extraction from business process models," Corvinus University of Budapest, Budapest, 2014.
[11] M. Arias, R. Saavedra, M. R. Marques, J. Munoz-Gama, and M. Sepúlveda, "Human resource allocation in business process management and process mining," *Manag. Decis.*, vol. 56, no. 2, pp. 376–405, Feb. 2018.
[12] D. Arena *et al.*, "Human resource optimisation through semantically enriched data," *Int. J. Prod. Res.*, no. December 2017, pp. 1–23, 2017.
[13] A.-K. Masum, L.-S. Beh, A.-K. Azad, and K. Hoque, "Intelligent Human Resource Information System (i-HRIS): A Holistic Decision Support Framework for HR Excellence," *Int. Arab J. Inf. Technol.*, vol. 15, no. 1, pp. 121–130, 2018.
[14] A. Singh *et al.*, "Method and System for auto-selection of Employee for training in an organization," US 2018/0039928 A1, 2018.
[15] A. Smirnov, A. Ponomarev, T. Levashova, and N. Shilov, "Ontology-based cloud platform for human-driven applications," in *2017 21st Conference of Open Innovations Association (FRUCT)*, 2017, pp. 304–310.
[16] N. Mittas, G. Kakarontzas, M. Bohlouli, L. Angelis, I. Stamelos, and M. Fathi, "ComProFITS: A web-based platform for human resources competence assessment," *IISA 2015 - 6th Int. Conf. Information, Intell. Syst. Appl.*, no. January, pp. 1–6, Jul. 2016.
[17] M. Bohlouli, F. Ansari, G. Kakarontzas, and L. Angelis, "An Adaptive Model for Competences Assessment of IT Professionals," in *Integrated Systems: Innovations and Applications*, Cham: Springer International Publishing, 2015, pp. 91–110.
[18] M. Bohlouli, N. Mittas, G. Kakarontzas, T. Theodosiou, L. Angelis, and M. Fathi, "Competence assessment as an expert system for human resource management: A mathematical approach," *Expert Syst. Appl.*, vol. 70, pp. 83–102, Mar. 2017.
[19] Z. Lili, "An Inverse Optimization Model for Human Resource Allocation Problem Considering Competency Disadvantage Structure," *Procedia Comput. Sci.*, vol. 112, pp. 1611–1622, 2017.
[20] T. Li *et al.*, "Human Resource Demand and Supply Model for National Research Institutions of China-," *Manage. Rev.*, vol. 4, no. 10, 2011.
[21] M. Xian-ying, "Application of Assignment Model in PE Human Resources Allocation," *Energy Procedia*, vol. 16, no. PART C, pp. 1720–1723, 2012.
[22] T.-H. Chien, Y.-I. Lin, and K.-W. Tien, "Agent-based negotiation mechanism for multi-project human resource allocation," *J. Ind. Prod. Eng.*, vol. 30, no. 8, pp. 518–527, Dec. 2013.
[23] Y. Miao, G. Jun, L. Xinggang, and Z. Huabo, "Staffing model for call center with customer patience variation," 2013.
[24] L. L. Zhang, X. N. Zhao, and Y. Y. Zang, "'Four-in-one' Human Capital Matching Method Based on Advantage Structure Identification[," *Syst. Eng. Pract.*, vol. 33, no. 8, pp. 2047–2056, 2013.
[25] K. B. B. Aviso *et al.*, "Allocating human resources in organizations operating under crisis conditions: A fuzzy input-output optimization modeling framework," *Resour. Conserv. Recycl.*, vol. 128, no. 8, pp. 250–258, Jan. 2018.
[26] Y. Li and J. Wang, "Marginal utility function based optimal human resource management model," *Xitong Gongcheng Lilun yu Shijian/System Eng. Theory Pract.*, vol. 36, no. 1, pp. 106–112, 2016.
[27] M. Arru, "PROKEX-K-Fit," *GitHub*, 2019. [Online]. Available: https://github.com/matteoarru/PROKEX-K-Fit.
[28] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *J. Mol. Biol.*, vol. 48, no. 3, pp. 443–453, Mar. 1970.
[29] S. Gillani and A. Ko, "Incremental Ontology Population and Enrichment Through Semantic-based Text Mining: An Application for IT Audit Domain," *Int. J. Semant. Web Inf. Syst.*, vol. 11, no. 3, pp. 44–66, Jul. 2015.
[30] A. Gábor and M. Arru, "Process Oriented Knowledge Transfer in the Public Administration," in *International Scientific-Practical Conference "Smart Government: Science and Technology,"* 2014, pp. 179–184.
[31] F. Forster, "The Idea behind Business Process Improvement: Toward a Business Process Improvement Pattern Framework," *BPTrends*, no. 1999, pp. 1–13, 2006.
[32] M. Armstrong, *Armstrong's handbook of human resource management practice*, 13th ed. Kogan Page Ltd., 2014.
[33] A. Gabor and M. Arru, "Knowledge Fit IES," *Mendeley Data*, vol. 1, Jul. 2018.
[34] E. W. Weisstein, "Unit Matrix," *MathWorld--A Wolfram Web Resource*, Wolfram Research, Inc., 2019.

A Comparative Study of Antminer+ and Decision Tree Classification Performances

CSABA BRUNNER
Corvinus University Budapest, MOL Group
eMail: csaba.brunner@uni-corvinus.hu

ABSTRACT

There is an increasing demand for knowledge discovery from large datasets. One of the more common extraction types is classification, placing entities into categories based on how similar their features are. Many new algorithms are experimented with, one example is antminer+. It is a swarm intelligence-based approach, modeling an ant colony's foraging behavior. A key question is, how well it performs on benchmark data compared to a known data mining algorithm. The answer depends on the classification metric of choice; it can achieve results comparable to other data mining algorithms.

Introduction

The data science community have always been at search for new algorithms that perform better one way or another. Some provided more accurate results, others achieved the same results faster than before. In this paper I will introduce and discuss a metaheuristic, more precisely an Ant Colony Optimization (ACO) based algorithm, called antminer+. According to [1] "A metaheuristic is a high-level problem-independent algorithmic framework that provides a set of guidelines or strategies to develop heuristic optimization algorithms". Metaheuristic algorithms are always heuristic in nature, this distinguishes them from exact algorithms. The latter was proven that given a finite (although often prohibitively large) time they will provide an optimal solution. On the contrary, metaheuristic algorithms were developed to provide a "good enough" solution in a "small enough" computing time. The most prominent metaheuristic algorithms are genetic algorithms, tabu search, simulated annealing and ant colony optimization. This article evaluates the possible application of ant colony optimization for data mining.

As described in [2]: "An artificial Ant Colony System is an algorithm based on agents that simulate the natural behavior of ants, developing mechanisms of cooperation and learning." Finding this solution is the emergent result of cooperation (between small agents called ants) and indirect communication via the environment. Antminer and the later antminer+ are ACO algorithms applied in complex data classification tasks.

Relatively few studies [3] have been carried out with the specific goal of comparing the classification performance of antminer/antminer+ and a widely applied data mining algorithm, for example, C4.5 decision trees. Such comparisons are necessary to determine the usefulness of new algorithms for data mining. First, a brief history of ant colony optimization in data science is introduced together with the key algorithmic steps of antminer. The improvements of antminer+ are then listed compared to antminer. Next, the applied methodology is introduced. This includes the dataset, the preprocessing steps, cross validation and the statistical testing methodologies applied. The final step is the discussion of the achieved results. The main question of this article is whether antminer+ can produce comparable results to a selection of decision tree methods in terms of the most common measures describing classification accuracy.

The antminer/antminer+ algorithms

Antminer and antminer+ algorithms were described in [2]–[6]. The idea of ant colony optimization is essentially based on the observed behavior of real-life ants. Ants communicate with each other not directly, but indirectly through their environment using

pheromone. Initially, an ant takes a random direction, leaving a trail of pheromone on its path. When it finds food, it returns to the nest. Each ant has a small preference for paths with stronger pheromone trail, therefore the ant returning the earliest will have the strongest pheromone trail and its path is more likely to be followed by the others, creating a positive feedback loop. The pheromone levels on other paths evaporate. This behavior is referred to as stigmergy [6].

The same principle is applied in ant colony optimization. Several artificial ants move through an environment incrementally constructing a solution by increasing the pheromone levels of the path they took. Each ant applies a stochastic local decision policy based on the pheromone level and a heuristic value. The heuristic is a problem-, while the pheromone is a history-dependent parameter. The path with the highest combined pheromone and heuristic values is taken. Once an ant arrives at its destination, the solution represented by its complete path is evaluated, and the pheromone trails of all possible paths are updated (increased for the path taken, else evaporated). The general steps in setting up a problem for ant colony optimization (including antminer and antminer+):

- Set up an environment that represents the problem domain
- A problem dependent heuristic evaluation function (η), providing a quality measurement of the solution
- A pheromone update rule, which considers both evaporation and reinforcement
- A probabilistic transition rule based on the heuristic function and the strength of the pheromone trail (τ) that determines the path taken by the ants
- And a convergence criterion which determines when to terminate the algorithm

[2], [5] reported the first application of ACO in classification on purely categorical data. The algorithm creates simple rules of *if antecedent, then consequent* structure, where the antecedent is a conjunction of simple rules. Antminer is a sequential covering algorithm; at every iteration, the subset of data covered by the constructed rule is excluded from the next iteration. The loop ends when a predefined stopping criterion is met. The environment is set up as a graph where each node represents a possible value for a given attribute.

The antminer algorithm is described in detail in [6]. The input parameters to control the metaheuristic behavior of the algorithm are:

- Max_Uncovered_cases: the maximum number of data rows that are not covered by any rule. Used as a stopping criterion in the outer loop of antminer.
- |Ants|: the number of ants (the number of potential rule candidates evaluated).
- Min_cases_per_rule: the smallest allowed number of records per rule, used to prevent overfitting.
- Converg_rules: if an ant constructed the same rule *Converg_rules* number of times consecutively, the algorithm concludes that the ants converged on the same path.

The environment consists of start and end nodes being the source and destination for each ant. The rest of the nodes stand for each unique value of every attribute and are fully connected except with nodes representing distinct values of the same attribute. During each iteration an ant starts from the start node with an empty rule, and decides on which edge to follow, iteratively constructing a new rule as a path in the environment.

Once the environment has been set up, the problem dependent heuristic can be calculated based on the Shannon entropy and information gain ratio, similarly to decision trees. The exact formulas for calculating entropy and information gain, according to [6], are:

$$Info(T_{ij}) = -\sum_{\omega=1}^{c}\left(P(\omega|T_{ij}) * \log_2 P(\omega|T_{ij})\right) \quad (1)$$

$$\eta_{ij} = \frac{\log_2(c) - Info(T_{ij})}{\sum_{i=1}^{m} x_i \sum_{j=1}^{p_i}(\log_2(c) - Info(T_{ij}))} \quad (2)$$

where c is the number of class values, T_{ij} is the subsample of the available data where attribute i equals value j and x_i is a constraint value to prevent picking two mutually exclusive attribute values in the same rule.

The other part of the probabilistic choice is the pheromone. It is initialized at the first iteration, then is kept updated later on. The pheromone initialization and update logics are [6]:

$$\tau_{ij}(t) = \frac{1}{\sum_{i=1}^{m} |value_i|} \quad (3)$$

$$\tau_{ij}(t+1) = \frac{\tau_{ij}(t)+\tau_{ij}(t)*Q}{\sum_{\forall ij \in rule} \tau_{ij}(t)} \quad (4)$$

where $|value_i|$ is the number of possible values for attribute i, whereas Q is the quality of the current rule. For antminer it is the product of sensitivity and specificity [6]:

$$Q = \left(\frac{TP}{TP+FN}\right) * \left(\frac{TN}{FP+TN}\right) \quad (5)$$

where TP is the number of true positives, FN are the false negatives, TN stands for the true negatives and FP for the false positives, all according to the confusion matrix of a rule (Table 1).

Table 1: The confusion matrix. Source: [7]

	Predicted class		
Actual class		P	N
	P	TP	FN
	N	FP	TN

The quality measure is used to guide the pheromone update calculation for attribute values selected by an ant. For every other value, evaporation is simulated by normalizing the previous pheromone levels by the sum of all (τ_{ij}) of the previous iteration.

Once both the heuristic values (η_{ij}) and the pheromones levels (τ_{ij}) per attribute value are available, the ant decides on the path to follow. This is performed probabilistically, according to the following calculation [6]:

$$P_{ij}(t) = \frac{\tau_{ij}(t)*\eta_{ij}}{\sum_{k=1}^{m} x_k \sum_{l=1}^{p_k} (\tau_{kl}(t)*\eta_{kl})} \quad (6)$$

Just like with the entropy calculation, the authors defined a constraint (x_k) value to avoid including contradictory attribute values in the final rule.

The antminer algorithm is executed in two loops. The first one runs until the number of data elements fall below a predefined threshold (Max Uncovered Cases). Within the loop, the environment and the starting pheromone levels are initialized. The second loop is where an ant constructs a rule. An ant keeps adding attribute values to its rule until all attributes were used in construction, or when the number of data entries covered by the rule would be fewer than a pre-defined threshold (Min cases Per rule), the latter criteria is used to prevent the overfitting of a rule. Rule construction, and the probabilistic choice between rules are controlled according to formulas (1)-(6). Overfitting is further avoided in a rule pruning step, where each attribute value is removed from the rule, and the resulting sub-rule qualities are evaluated. If a sub-rule quality is better than the original, the rule gets replaced. In the final steps of the second iteration, the pheromone levels are updated, and the rules are evaluated for convergence. If either the number of ants (candidate rules) reaches a limit (|Ants|) or a given rule was constructed Converg_rules number of times, the second iteration stops. At this point some additional operations are carried out within the first loop: the best rule is selected from a list of rules constructed by all ants, appended to a final list of discovered rules, and all the data elements covered by the best rule are removed from the data.

Antminer was the first, but not the only application of ACO in data mining. The later iterations were antminer2 [8], antminer3 [9] and antminer+ [3], [4]. Antminer2 improved on antminer by using a density estimation function for the heuristic value. Antminer3 provided two key improvements over previous versions: a different pheromone update function was defined, one that relied on the sum of sensitivity and specificity, rather than the product, and a new rule construction process, one that favored attributes unused in previous iterations of the algorithm.

However, the biggest improvements were brought up with antminer+. Unlike antminer, antminer+ defined the environment as a directed acyclic graph (DAG). This new environment causes a more efficient choice of attribute value, as an ant now only decides about the value of the next attribute, instead of all attributes. Moreover, antminer+ supports ordinal and (as an extension) numeric attributes. This is performed by transforming the value nodes of a given ordinal attribute. For ordinal attributes, two lists of nodes are created, one for the lower bound choice and one for the upper bound, constrained by

the lower bound. When a rule for an ordinal attribute is generated, then this is interpreted as the value must be between the lower and upper bounds. The environment also allows weight parameters α and β for pheromone levels and heuristic values to be added, controlling which one of the two to be favored more during rule construction. Finally, all class values, except the default majority class, are included in the environment, allowing multiclass classification to be performed. Other improvements over antminer include the application of the MAX-MIN ant system proposed by [10], new heuristic and pheromone update calculations and early stopping.

A Matlab [11] implementation of antminer+ is available in [4] which was used as part of this research. Moreover, an extensive optimization study of the algorithm has been carried out in [3].

Decision trees

The performance of antminer+ on its own, however, is not enough. For this reason, a decision was made to compare the results it achieved with a widely accepted algorithm in the data science community, namely decision trees. Altogether, the measurements of antminer+ were compared with three decision trees:

- C4.5, based on [12]
- CART, based on [13]
- And Conditional Inference trees, based on [14]

The most prominent difference between the three algorithms can be found in their splitting criterion (Gain ratio for C4.5, Gini index for CART and statistical permutation testing for CI trees). A detailed introduction of each decision tree algorithm is not a goal for this paper, more information on them can be found in their respective papers.

Classification performance

The previous section briefly touched the topic of classification performance with the inclusion of the confusion matrix, seen in Table 1. There, sensitivity and specificity were calculated per rule, however they can be used to describe the results of the whole algorithm as well. Some more of the most common performance metrics are accuracy, precision and F_1-score. The formula for each is, according to [7]:

$$accuracy = \frac{TP+}{TP+FP+TN+FN} \quad (7)$$

$$recall = \frac{TP}{TP+FN} \quad (8)$$

$$precision = \frac{TP}{TP+FP} \quad (9)$$

$$F_1\ score = \frac{2*precision*recall}{precision+recall} \quad (10)$$

Further methodology and data collection are based on calculations (7)-(10).

Methodology

This section will discuss the comparison methodology between antminer+ and decision trees. With this comparison, the article attempts to determine whether antminer+ yields similar results, and whether it is better in terms of classification performance. First, a suitable dataset had to be found. One such dataset is the Hepatitis dataset provided by [15] containing 155 cases of 19 attributes. The task is to predict the outcome of Hepatitis infections with respect to demographic, chemical and pathological features. This dataset could not be used for training or testing in its initial state, due to incorrect attribute representations and the existence of missing values. The former was handled by recoding the attributes from numerical to categorical values. For missing values, two strategies were applied: attributes which had missing values in more than 40% of the cases were removed entirely from the dataset. The attributes having less than 40% missing value rate were imputed with either their mode (in case of categorial attributes) or their median (in case of numerical attributes).

The preprocessed dataset was then divided into a 60% training and a 40% test sample using stratified split. Each of the algorithms were trained in 10 times repeated 10-fold cross validation on the training set. Though available, the results of this cross validation were not used later in the results evaluation phase, as every training fold were used at least once for testing, increasing the likelihood of information leakage (information on the dataset declared for testing leaks into the training phase). Instead, only the models (the rule lists constructed by antminer+ and each decision tree algorithm) from the cross-validation steps were inherited and tested on the remaining 40% sample.

The training process for antminer+ was executed first. To avoid differences (likely caused by the randomness of the data fold generation) between the algorithm executions, the training and test folds were saved during the cross validation of antminer+ and were later reused during the decision tree constructions. Moreover, antminer+ has a variety of parameter settings. These settings were described in detail in [3]. Only the settings impacting model performance the most were used, namely:

- A limit of 100 was imposed on the iteration count, though studies shown that much less was enough.
- The remaining dataset size limit was set to 15% of the original.
- The heuristic calculation was the same as in [4], despite more being available.
- The implementation also contains strategies for attribute selection. These are controlled by two parameters. The first determines whether attribute selection could be used in the algorithm at all, and if yes, then what calculation should it be based on. To maintain similarity to the most popular decision tree algorithm (C4.5), the experiment used gain ratio. The maximum allowed number of attributes selected was set to the default 10.
- The rest of the parameters used their respective default values proposed by [3], [4]. One example is the number of ants, which remained at 1000.

Like antminer+, though to a lesser extent, each of the decision tree algorithms also used customized parameter settings. For C4.5, The number of observations per leaf were minimized at 10. If a split were to result in two nodes with less observations than that, the leaves were not split any further. A similar setting was applied to CART as well, though there the number had to be at least 5. As for conditional inference trees, all settings remained the default.

The results from the final validation are then evaluated by comparing the empirical distributions of classifier performance metrics derived from the confusion matrix discussed previously. The empirical distributions of these metrics were then evaluated graphically using empirical cumulative distribution functions (or ECDFs) [16], and statistically by two-sample Kolmogorov-Smirnov hypothesis tests for comparing distributions [17].

Another evaluation was performed between the global average classifier performance values. This was based on a more traditional comparison of performance metric means with 95% confidence interval approximated using standard deviations. These aggregated results were also compared with a public ranking of the Hepatitis dataset found at [18], available for accuracy only.

Both evaluation processes involved some level of aggregation. The graphical/statistical evaluation needed per model performance metrics, which was available for accuracy only. The rest were macro averaged to a repeat + fold level, excluding per class granularity. The global average calculations, as the name suggests, aggregated the entire range of values for all metrics for both average and standard deviation calculations.

Two programming languages were used in this experiment: the antminer+ algorithm was only implemented in Matlab, hence the application of Matlab. Data preprocessing, stratified split, decision tree model training and results evaluation were implemented in KNIME, an open-source data manipulation software based on the Java programming language and the Eclipse Integrated Development Environment. More information on KNIME is available in [19].

Results

This section discusses the classification performance evaluation by comparing antminer+ and decision tree results.

The first experiment with the resulting data consisted of plotting ECDFs of the selected performance metrics and performing Kolmogorov-Smirnov tests. The two are interconnected; the Kolmogorov-Smirnov test performs a statistical comparison between two numeric samples based on their distribution of values. An ECDF captures this distribution graphically. In fact, the Kolmogorov-Smirnov test calculates the maximum distance between two ECDFs.

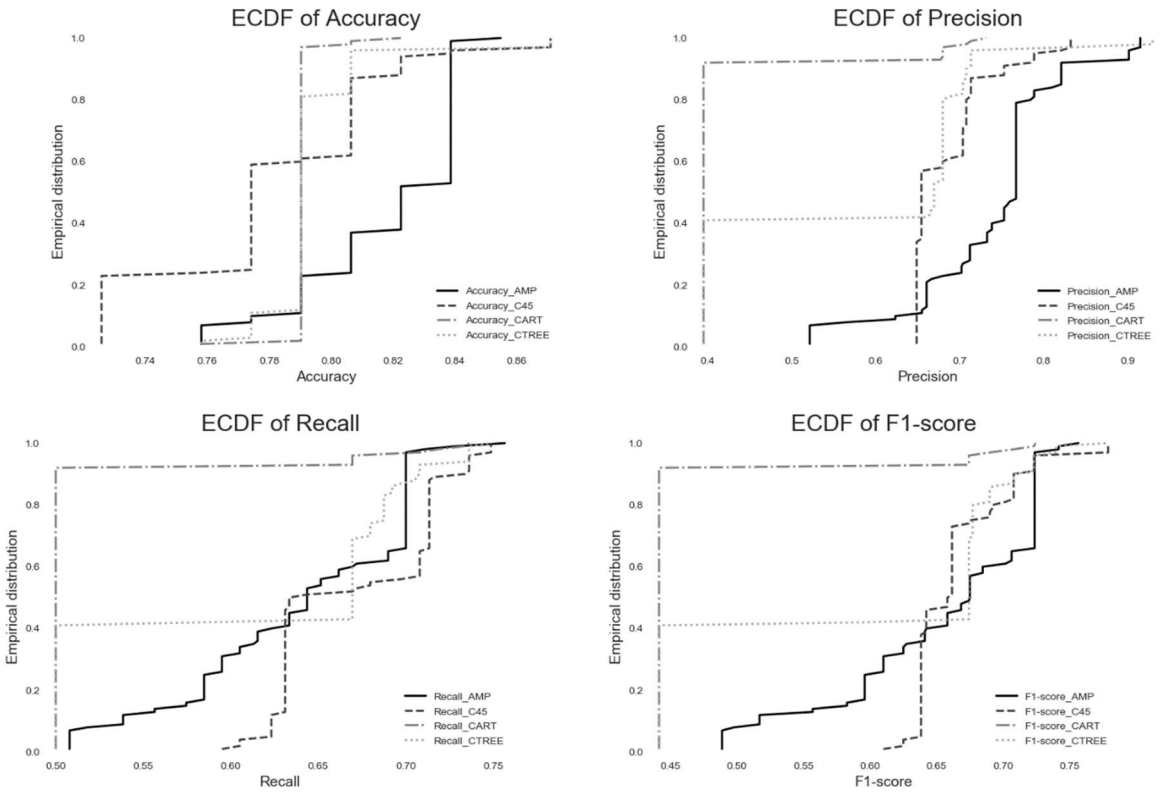

Figure 1. Empirical cumulative distribution functions for each metric

The ECDFs corresponding to the selected performance metrics can be seen on Figure 1. Each of the lines represent the antminer+ and decision tree models. ECDFs visualize the observed distribution of a statistical variable, in this case the collected measurements. On average, the closer a line is to the right edge of the plot, the better the corresponding model performs. Ctree and CART performed the worst out of the four with regards to precision, recall and F_1-score. These graphical representations, though demonstrative, are difficult to draw exact conclusions from. Therefore, the Kolmogorov-Smirnov tests were calculated too, provided in Table 1.

The Kolmogorov-Smirnov tests were calculated for each metric between the results of antminer+ and each decision tree model. The null hypothesis of the test states that two samples are from the same statistical distribution. This must be rejected at all common confidence levels for all the evaluated metrics; antminer+ provided significantly different results than any of the decision trees.

Table 1. Kolmogorov-Smirnov test results

Metric	Model	Reject H0	Statistic	p-Value
Accuracy	C4.5	true	0.62	0
	CART	true	0.85	0
	CTREE	true	0.71	0
Precision	C4.5	true	0.54	4.34E-13
	CART	true	0.92	0
	CTREE	true	0.63	0
Recall	C4.5	true	0.41	1.00E-07
	CART	true	0.92	0
	CTREE	true	0.41	1.00E-07
F_1-score	C4.5	true	0.32	7.14E-05
	CART	true	0.92	0
	CTREE	true	0.41	1.00E-07

Table 2. Aggregated Performance metrics with 95% confidence intervals

Metric	Antminer+ (mean ± 2stdev)	C4.5 (mean ± 2stdev)	CART (mean ± 2stdev)	Ctree (mean ± 2stdev)
Accuracy	**0.818(+/-0.051)**	0.780 (+/-0.075)	0.791 (+/-0.010)	0.794 (+/-0.037)
Precision	**0.739(+/-0.180)**	0.686 (+/-0.100)	0.419 (+/-0.163)	0.575 (+/-0.315)
Recall	0.638 (+/-0.128)	**0.671 (+/-0.092)**	0.516 (+/-0.109)	0.609 (+/-0.188)
F_1-score	0.654 (+/-0.151)	**0.665 (+/-0.073)**	0.462 (+/-0.138)	0.587 (+/-0.247)

Table 3. Hepatitis classification performance. Source: [18]

Rank	Method	Accuracy %	Reference
1	Weighted 9-NN	92.9	[20]
2	18-NN, stand. Manhattan	90.2±0.7	[20]
3	FSM with rotations	89.7	Rafał Adamczak, in [18]
4	15-NN, stand. Euclidean	89.0±0.5	[20]
5	VSS 4 neurons, 5 it	86.5±8.8	WD/MK, in [18]
6	FSM without rotations	88.5	Rafał Adamczak, in [18]
7	LDA, linear discriminant analysis	86.4	[21]
8	Naive Bayes and Semi-NB	86.3	[21]
9	IncNet	86	Norbert Jankowski, in [18]
10	QDA, quadratic discriminant analysis	85.8	[21]
11	1-NN	85.3±5.4	[21]
12	VSS 2 neurons, 5 it	85.1±7.4	WD/MK, in [18]
13	ASR	85	[21]
14	Fisher discriminant analysis	84.5	[21]
15	LVQ	83.2	[21]
16	CART (decision tree)	82.7	[21]
17	MLP with BP	82.1	[21]
18	ASI	82	[21]
19	LFC	81.9	[21]
20	Antminer+	81.8±2.55	Own research
21	Decision tree (Ctree)	79.4±1.85	Own research
22	Decision tree (CART)	79.1±0.5	Own research
23	RBF (Tooldiag)	79	Rafał Adamczak, in [18]
24	Decision tree (C4.5)	78±3.75	Own research
25	MLP+BP (Tooldiag)	77.4	Rafał Adamczak, in [18]

However, the Kolmogorov-Smirnov test is only useful for assessing the difference between two samples, and not for determining whether a model performs better than the other on average. To determine this, the means and standard deviations were calculated for each metric and model. The standard deviations were then used to approximate the 95% confidence intervals. Table 2 contains the achieved the results. Antminer+ provided the best accuracy and precision values among the 4 models, whereas C4.5 provided the best recall and F_1-score.

The remaining decision trees, CART and ctree tied for second best for accuracy, however they struggled with achieving high precision, recall and F_1-score values. This was caused by an imbalance between the positive and negative classes in the Hepatitis dataset, introducing a bias towards the majority class. The high accuracy levels for both the model and results from Figure 1 support this observation, the former providing a good example why accuracy alone should not be trusted. A third possibility is to compare model performance to external benchmarks. [18] collected cross-validation accuracies over several datasets and scientific papers, including the Hepatitis dataset.

The external results can be seen in Table 3. The author attempted to collect results obtained using 10-fold cross validation wherever it was possible. The accuracies in (

) were all averaged with their respective standard deviations. Antminer+ performed the best, reaching rank 20, while ctree, CART and C4.5 each ranked 21, 22 and 24. The authors of [18] stated that the Hepatitis dataset had little variation in its results, pointing out the missing value handling as a potential area of improvement.

Conclusion

Depending on the model performance metric, antminer+ can achieve similar or better results compared to other data mining algorithms, like decision trees. The methodology of performing nonparametric tests and performance metric aggregations on multiple cross-validation results has proven to be a robust method for comparing classifier models. The possible reason for a lack of widespread usage of antminer+ may lie in other factors, for example, the complexity of the more popular data mining algorithms does not increase in an NP complete manner with regards to observation count, or other methods exist to reduce the number of observations, simplifying model training. measurement and comparison of execution times is the top candidate for future research. Another opportunity for improvement is the parameter optimization of antminer+ with the help of measurements and external benchmarks.

References

[1] K. Sörensen and F. W. Glover, "Metaheuristics," in *Encyclopedia of operations research and management science*, Springer, 2013, pp. 960–970.

[2] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, "An ant colony based system for data mining: applications to medical data," *Proc. Genet. Evol. Comput. Conf.*, pp. 791–797, 2001.

[3] B. Minnaert, D. Martens, M. De Backer, and B. Baesens, "To tune or not to tune: rule evaluation for metaheuristic-based sequential covering algorithms," *Data Min. Knowl. Discov.*, vol. 29, no. 1, pp. 237–272, 2013.

[4] D. Martens *et al.*, "Classification With Ant Colony Optimization," *IEEE Trans. Evol. Comput.*, vol. 11, no. 5, pp. 651–665, 2007.

[5] R. Parpinelli, H. Lopes, and A. A. Freitas, "Data Mining with an Ant Colony Optimization Algorithm," *IEEE Trans. Evol. Comput.*, vol. 6, no. 4, pp. 321–332, 2002.

[6] D. Martens, B. Baesens, and T. Fawcett, "Editorial survey: Swarm intelligence for data mining," *Mach. Learn.*, vol. 82, no. 1, pp. 1–42, 2011.

[7] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*. Elsevier Ltd, 2011.

[8] B. Liu, H. A. Abbass, and B. McKay, "Density-based heuristic for rule discovery with ant-miner," in *The 6th Australia-Japan joint workshop on intelligent and evolutionary system*, 2002, vol. 184.

[9] B. Liu, H. A. Abbas, and B. McKay, "Classification rule discovery with ant colony optimization," in *Intelligent Agent Technology, 2003. IAT 2003. IEEE/WIC International Conference on*, 2003, pp. 83–88.

[10] T. Stützle and H. H. Hoos, "MAX--MIN ant system," *Futur. Gener. Comput. Syst.*, vol. 16, no. 8, pp. 889–914, 2000.

[11] MathWorks, "MATLAB and Statistics Toolbox." MathWorks, 2018.

[12] J. R. Quinlan, *C4.5: programs for machine learning*. Elsevier Ltd, 1993.

[13] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, 1st ed. New York: Routledge, 1984.

[14] T. Hothorn, K. Hornik, and A. Zeileis, "Unbiased recursive partitioning: A conditional inference framework," *J. Comput. Graph. Stat.*, vol. 15, no. 3, pp. 651–674, 2006.
[15] P. Diaconis and B. Efron, "Computer intensive methods in statistics," *Scientific American*, vol. 248, no. 5. pp. 115–130, 1983.
[16] E. F. Drion and Others, "Some distribution-free tests for the difference between two empirical cumulative distribution functions," *Ann. Math. Stat.*, vol. 23, no. 4, pp. 563–574, 1952.
[17] G. W. Corder and D. I. Foreman, *Nonparametric statistics: A step-by-step approach*. John Wiley & Sons, 2014.
[18] W. Duch, R. Adamczak, and K. Grudziński, "Datasets used for classification: comparison of results," 2012. [Online]. Available: http://www.is.umk.pl/~duch/projects/projects/datasets.html. [Accessed: 17-Jan-2019
[19] M. R. Berthold *et al.*, "KNIME-the Konstanz information miner: version 2.0 and beyond," *AcM SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 26–31, 2009.
[20] W. Duch, K. Grudzinski, and G. H. F. Diercksen, "Minimal distance neural methods," in *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on*, 1998, vol. 2, pp. 1299–1304.
[21] B. Šter and A. Dobnikar, "Neural networks in medical diagnosis: Comparison with other methods," in *International Conference on Engineering Applications of Neural Networks*, 1996, pp. 427–430.

Solving Exercise Generation Problems Using the Improved EGAL Metaheuristic Algorithm

BLANKA LÁNG
Corvinus University Budapest
eMail: blanka.lang@uni-corvinus.hu

ABSTRACT

Exercise generation problem is a well-known problem worth to investigate. This paper presents how to apply a previously developed diversity-oriented harmony search metaheuristic algorithm (EGAL) for different exercise generation problems, in those cases when the nature of the problems and the examined subjects are very different. The necessary modifications in the algorithm, what we have to do when improved EGAL is applied for these different tasks are summarized. The behaviors of the tasks for which the improved EGAL algorithm can be applied effectively, are summarized as well. Finally, some test results are given for some widely accepted benchmark functions.

Introduction

In this paper an improved algorithm, and different exercise generation problems will be presented. The reader can find the related works, detailed mathematical background, and the description of the optimization search problem in our former publication [1]. The harmony search algorithm and the EGAL algorithm (Exercise Generator Algorithm) are described below.

EGAL is a type of harmony search algorithm (HS). HS is based on an analogy with music improvisation process, where the musicians improvise the sounds to obtain better harmony. It was developed in 2005 by Lee and Geem [2]. The algorithm converges to a global optimum due to its population

based nature. In the harmony memory, (called *repertoire* in the language of music) there are individuals, or vectors (called *melodies* in music). The aesthetic value of a melody is represented by the fitness function value. The higher this fitness function value is, the higher the quality of the melody is. Using improvisations we would like to minimize or maximize the value of this function, to find the global optimum (or in the language of music: to find the best harmony). A HS melody is a vector, in our case a string of binary cells. If the i^{th} cell has value 1, than X_i is selected, otherwise not selected. Each solution is evaluated with respect to their objective functions.

First of all, the algorithm parameters that are required to solve the optimization problem are specified. We select the decision variables in such a way to obtain the best fitness value results. The first important parameter is the harmony memory size. The improvisation process is driven by two parameters – they are probability values between 0 and 1; *HMCR* (harmony consideration rate), *PAR* (pitch adjusting rate). To select the best values depends on the given problem. We will specify these parameter values in each problem cases.

After selecting these values, we have to find other four parameter values. The first three are the maximum generation number (*MaxGen*), the size of the population (*PopSize*), and the number of the possible questions or conditions or functions (*PSize*).

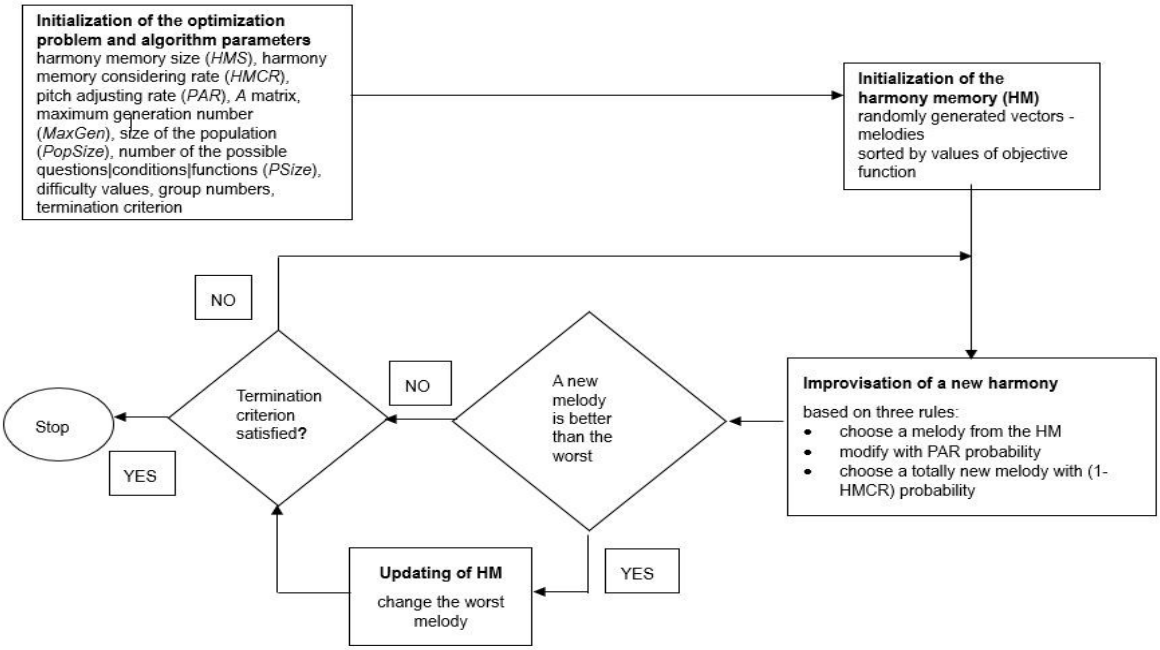

Figure 1: The EGAL algorithm

When we use EGAL for exercise generation problems, we would like to find a certain "best" subset of predefined questions. It will be clarified later, what does it mean exactly: "best". We have to note, that sometimes there are not questions in the task, but for example we would like to select some conditions from a predefined condition set or functions from a predefined function set. One has to generate tests for a class of students where tests consist of a set of questions, meaning that the *PopSize* equals 30 in the case of the all problems. The *PopSize* greatly can be increased, we tested these problems for example in the case of *PopSize* equals to 100.

After these decisions, we have to create a quadratic weight matrix (*A*). The size of this matrix equals *PSize*, it consists of numbers between 0 and 1, where each a_{ij} in the matrix shows a certain preference. This a_{ij} value equals 0, if the teacher does not

want to select the i^{th} and j^{th} questions in the same time, and equals 1, if the teacher especially wants to select them together in the same task. The larger value between 0 and 1 means that we would like to select the questions in the same time preferably. Specifying these a_{ij} values is slightly subjective in most cases, it depends on the teacher's decision and the behavior of the problem.

After these initializations, we have to decide how many questions do we want to ask from a student in one test. This value (*TSize*) is obviously less than *Psize*. It was set 6 in the case of the all problems.

1 0 0 0 0 1 0 0 1 0 0 0 0 0 1 0 0 1 0 0 0 0 0 1

Figure 2: An individual in the population

After specifying the parameters we have to initialize the first population. The first population is filled with randomly generated vectors. In each vector has exactly *TSize* 1 bits, and the others are zero bits. The 1 value means, that we would like to select the i^{th} question in this task, zero means that we do not want to select that.

The next step is to calculate the fitness function. At the time when the EGAL algorithm was developed, we wanted to focus simultaneously on the diversity measure and fitness value. Our intention was to maximize the fitness values and the distance between individuals in the population in the same time. First, we calculated the initial fitness values according to the values of the *A* matrix and the population vector values.

Additionally, the diversity measure was integrated into the fitness value. Diversity measure was calculated in the following way. Let us calculate first the distance for two individuals (*t, p*) from the same population:

$$d(h,k) = \sum_{i=0}^{PSize-1} |h[i] - k[i]|$$

Then let us calculate the diversity measure for the *p* individual

$$D(p) = \sum_{t=0}^{PopSize-1} d(p,t)$$

The selection function combines different values by a single operator ○ depending on the diversity measure used for the task. Function ○ could be for example an addition, subtraction or a multiplication/division. In this paper, the ○ is simply the addition operation. The fitness and diversity values may be in different range of numbers, thus first we have to standardize them. After the standardization let us calculate the fitness value for the *p* individual (*F(p)*) in the following way:

$$F(p) = F_{original}\ (p) + D(p)$$

After this calculation the individuals are sorted by the values of the objective function. When we use this combined fitness function, we will focus simultaneously both on diversity and fitness.

Then the new melody improvisation phase follows. According to the harmony consideration rate (*HMCR*) we choose a melody from the memory, or we generate a totally new, random melody with probability 1-*HMCR*. According to the pitch adjusting rate (*PAR*), the melody, selected from the memory, is modified with a probability *PAR*. Afterwards, if the new melody is better than the worst in the population, the worst one will be replaced by the better new one, otherwise we through away the new one. A new melody is better for us than the old one, if its fitness value lower/higher than the old one's value. This melody improvisation phase is repeated until the termination criterion is met. According to the applied replacement strategy, the quality of the population is increasing step by step. In this publication the termination criterion was the following: the algorithm was stopped if in the last 10 populations the increasing of the average fitness value was less than ε, ε=0.0000000001. At the end of the process we get different tasks for a class of students, where all the tests have good fitness function value.

Considering group number

Let us take the css task generation problem mentioned in this previous publication [1]. Our trade-off goal was complex: the exercises should cover the more course material, and the tasks has to be as different as possible in the same time.

In this case, every student gets a sample page

as an image, the task is for them to answer for some simple yes-no questions: "Can you see the following css styles in the picture: *h1 {font-style:italic;}, h1 {letter-spacing:20px;}, body{ background-color: yellow;}, hr {width:20%;}*, etc".

Let us see the *A* matrix in this case. Naturally there are certain questions we do not want to ask together due to logical reasons, for example using two or more different colors or different alignments for the same tag, because the subsequent style overwrites the former style. We have to write zero values in the *A* matrix to the appropriate places. And there are certain questions we do not want to ask together so much for pedagogical reasons, for example using the exactly same style for different tags, because in this case the result covers less the syllabus. In this case we have to write a value less than one, but bigger than zero in the *A* matrix to the appropriate places. To determine the all values in the *A* matrix depends on the teacher's sometimes subjective decisions.

We have to keep in mind that EGAL is not suitable for solving every kind of task generation problems. We summarized the characteristics of the solvable problems when the original EGAL can be applied effectively: (1) We would like to find more subsets of predefined questions (or conditions, or functions etc.) (2) If we check two possible solutions, we can decide clearly which is the better solution for us according to the fitness value. (3) We would like different individuals in a population, according to the suitable diversity measure, which was integrated into the fitness function. (4) There are too many possible solutions, we cannot consider all of them. (5) We do not want to find an exact optimum, we are satisfied with a good enough set of solutions if those are "enough close" to the optimum.

In the cases when we would like to solve more complicated task generation problems it is necessary to modify, to improve the original EGAL, to integrate new criterion into the fitness function. Here we can use the improved EGAL for more complex problems. In the followings some new criteria will be presented, the required changes in the algorithm will be determined, and the characteristics of these more complex problems will be summarized.

EGAL algorithm solved effectively and fast the css task generation problem [1], but a new integrated criterion can result better tasks. This new criterion was integrated into the fitness function: the "group number". On the picture the heading (H1 tag), a paragraph (p tag), and a list (ol tag) are formatted using css styles.

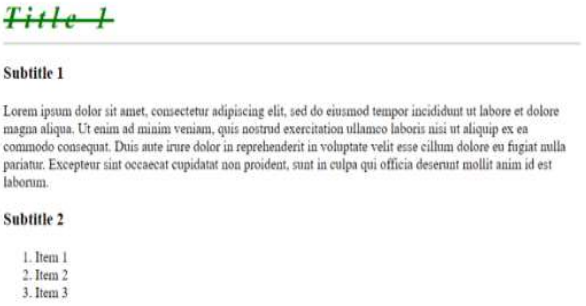

Figure 3/A Original EGAL's sample run for the css generation problem: a bad and a good result

Figure 3/B Original EGAL's sample run for the css generation problem: a bad and a good result

In the paper [1] we used the original EGAL algorithm, and that solution was more or less satisfying: the result covered enough the course material, and asked it in different ways. The algorithm was fast enough as well. But the original algorithm did not handle the "group problem." The exercise is obviously better for us, covers more of the course material, if we select not only a few, but more tags to format. According to the original algorithm, in extreme cases, we can get a task in which only one tag is formatted. For example, if the all six questions belong to only for the first heading tag (H1), this task has a worse

quality, but if the questions belongs to two or three groups, than the task has better quality, i.e. it covers the material better. On the Figure 3/A we can see an extreme task, in which we formatted only the heading tag, and a better example on the right, in which we formatted all of the three tags (h1, p, ol tags).

The original EGAL algorithm protects us to some extents from these kind of extremely bad quality solutions – due to the random nature of the HS -, but if we want to be sure, that the algorithm rewards when we formatted more tags, an extra criterion is required. For this purpose a "group number" (denoted by *m*) was calculated and integrated into the fitness function. We do not require formatting all tags in every case – because this strongly restricts the set of the possibly solutions – but we wanted to reward the result if the questions were from more groups. If the questions were from exactly two groups in the case of *p* individual, an *m(p)*=1.5 multiplier was integrated into the fitness function, if the questions were from exactly three groups (the heading tag, the paragraph tag, and list tag were formatted) an *m(p)*=1.9 multiplier was integrated into the fitness function. Note, that specifying the multipliers are subjective decisions, the teacher choose these values considering his/her demand and the task behavior. In this case the fitness function for the individual *p* will be:

$$F(p) = (F_{original}\ (p) + D(p)\ ) * m(p)$$

Obviously, if we would like to see the improvement of the effectiveness, we have to analyse the old and the new results. In order to compare the effectiveness, the group number is investigated. Comparing the group numbers, we can see, that the average group number value increased by 24.7 %. This result was computed as the average of 20 runs. The result coincides with our intent. The improved algorithm asks more different question groups, it covers the subject obviously better, the quality of the algorithm is evidently higher.

Summarizing the characteristics of the solvable problems where this improved EGAL with group number can be applied effectively, we can define: (1) The above characteristic list is also right in this case. (2) The questions can be classified into some groups, and we would like to select questions from more groups, but not always compulsory from every groups. (3) We can determine suitable values to reward the number of the used groups.

Considering difficulty

Let us take an other task generation problem. In this case we have to modify the original EGAL algorithm as well. This example is a form validation task generation case. In this case the student first has to create a web form, for example using the following input fields: 2 text boxes, a password box, 3 check boxes, 3 radio buttons, and a number input field. In this case we do not want to create "questions" as in the previous example, but some "conditions" are selected from a predefined condition set. The students get these selected conditions, and have to write a program which satisfies the condition set below.

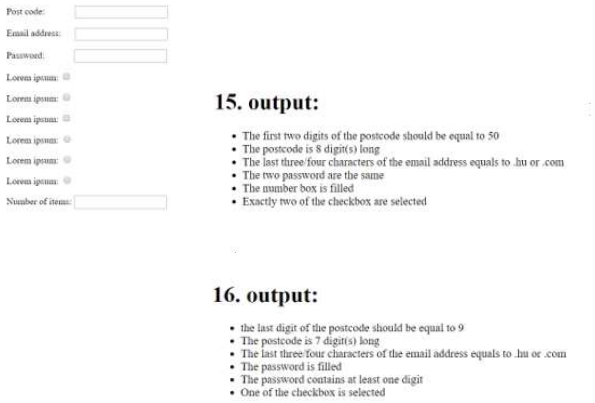

Figure 4. Improved EGAL's sample run for the form validation task generation problem using difficulty

This problem is different from the first example: in this case the difficulty of the conditions are diverse. For example, this is a very easy condition: "the postcode is to be filled", but this other one is much more difficult: "two of the checkboxes are selected at most". It would be unfair if one student would have a set of easy conditions, but the other would get a set of difficult conditions.

In this case a difficulty criterion is integrated into the algorithm. At the start of the process, the teacher has to fill a *difficulty[ ]* array. This array should be *PSize* long, and it contains integer values. These values represent the difficulty of the conditions,

between a minimum (*DiffMin*) and a maximum (*DiffMax*) value. Note, that the selection of these minimum and maximum value depends on the teacher's decision. After these selections the teacher has to fix a *Difficulty* value.

$$Difficulty = \sum_{i=0}^{PSize-1} difficulty[i]$$

In this case we would like to select a set of conditions in according with this *Difficulty* value, so that the cumulated difficulty of the condition subset should be equal with this *Difficulty* value. Obviously on the one hand the *Difficulty* value should be bigger than the sum of *TSize* minimum difficulty value, and on the other hand should be smaller than the sum of *TSize* maximum *difficulty* value. It is not worth to select the *Difficulty* value close to the maximum and minimum limits, because in that case the selection would reduces the freedom of the process. On the other hand, to select an appropriate *Difficulty* value is not easy, for example, if every *difficulty* value is even, could not be given an odd *Difficulty* value as a sum.

Thereafter when the first population is initialized, only the appropriate individuals will be accepted, if its *difficulty* equal the *Difficulty* value. After, when a new individual improvisation phase follows, we will accept the new one, only if its difficulty value equals the given *Difficulty*. Let us calculate the fitness value for the individual *p* in the following way:

$$F(p) = F_{original}\ (p) + D(p)$$

In our case, first the *DiffMin* and the *DiffMax* values were set: 1 and 3. After the array was filled, and the *Difficulty* value was set: 12. You can see the result in the picture: all of the condition set has the same *Difficulty*, which means for us the task is "fair" for the students. The *Difficulty* can be bigger or less, if the teacher would like an other task which has a different Difficulty.

We note that later we can modify the difficulty values according to the students' results. For example if a condition, a question was very easy or very difficult, we can decrease or increase its original difficulty value.

Summarizing the characteristics of the solvable problems it can be stated:

(1) The above characteristic list is right in this case as well. (2) A list of conditions (or questions or functions), etc. is given. (3) The *difficulty* of these list items / conditions are different from each other, but we would like to select subsets which are equally difficult, which has the same *Difficulty*.

Considering group number and difficulty

Let us take the third, most complicated task generation problem. In this case we have to modify the original EGAL algorithm as well. This example is an Excel aggregating function task generation case. In this case two new criteria should be integrated into the fitness function: considering the "difficulty" and the "group number" together.

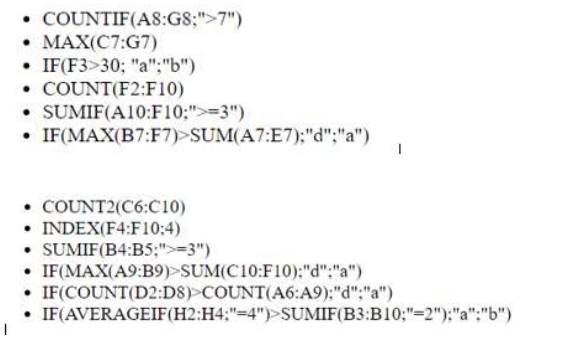

Figure 5. Two sample runs for the Excel function task generation problem using difficulty and group number: an easy task (Difficulty=20) and a difficult task (Difficulty=30)

In this case, every student gets two images. The first image is a simple excel table, the second image is a set of some Excel functions. Without using the Excel program, the students have to give the result of these functions. So, the algorithm should generate some sets of Excel functions. These sets should be different, pedagogically and logically "good", and can handle the group number and the difficulty problem, which are mentioned above. The first two conditions are satisfied by the original EGAL function, as was discussed above.

The difficulty problem: the difficulty of the Excel functions are very different. For example, this is a very easy function *"= SUM(B2:B5)"*, but this other one is much more difficult: *=IF(SUMIF (A2:F2;"= 4";B3:D3) >=AVERAGEIF (B5:F5;"=2";B6:D6);"a";"b")*. It would be unfair if one student got a set of easy

conditions, and the other student got a set of difficult conditions. In this case the *difficulty* criterion is integrated into the algorithm again, as above was discussed. The *DiffMin* and the *DiffMax* values were set: 1 and 10, and the array was filled. The *Difficulty* value was set: 20 and 30.

The group number problem: the Excel functions were classified into 3 groups for pedagogical reason. A "group number" was calculated and integrated into the fitness function again: the result was rewarded if the questions were from more groups using multipliers. You can see one result of the algorithm in the picture: all of the Excel function set has the same difficulty, thus the task is "fair" for the students.

Let us calculate the fitness value for the individual *p* in the following way:

$$F(p) = (F_{original}\ (p) + D(p)\ ) * m(p)$$

Summarizing the characteristics of the solvable Figure when this improved EGAL with group number and difficulty can be applied effectively: (1) The above characteristic list is right in this case as well. (2) The questions can be classified into some groups, and we would like to select questions from more groups, but not compulsory always from every groups. (3) We can determine suitable values to reward the number of the used groups. (4) A list of questions or conditions, etc is given. (5) The *difficulty* of these list items/considitons/functions are different from each other, but we would like to select subsets which are equally difficult.

Test results

It is very hard (almost impossible) to give perfect test results for the above exercise generation tasks. First of all, to select the maximum number of the iterations or the termination criterion strongly depends on the actual subject. Furthermore it does not depend on only the number of the questions, the algorithm parameters, but strongly on the nature of the problem, the weight matrix, selecting matrix elements, the actual groups (if there are), on the actual difficulty values (if there are), and on the mutual dependencies of these values. Summarizing, the given test results worth only for widely accepted benchmark functions. In order to see the differences between the algorithms we compared the HS algorithm and EGAL on the following problem settings.

In our former publication [1] we gave some running test results for the well known Six-Hump Camelback function [3] [4], which is a standard benchmark function for optimization problems. Since this function has two optima, the task was to find as many alternative solutions as possible. One of our conclusion was that EGAL is effectively focusing on finding a diverse set of solutions. The running time was obviously higher than the running time of the original harmony search algorithm, because HS can focus on finding a single optimum, so it converges faster, but that finds only a single optimum.

Now our intention was to give other test results. In this case these functions are well-known test functions to evaluate runtime performance: Rosenbrock function [4] [5], and Goldstein and Price function [6]. The first function is a "hard" benchmark function, due to the long narrow and curved valley present in the function. However, even though this valley is easy to find, to converge to the global minimum is difficult. This function has one global minimum: (1.0, 1.0), and the optimum function value is zero. The bounds were set between -10.0 and 10.0 for the two design variables, the number of iteration was set to 25000. We used the following settings according to [2]: *HMCR*=9, *PAR*=0.35, population size: 20, and 20 independent runs. In each tests, HS algorithm found a candidate around the optimum, the average best was: (0.9996, 0.9992) in approximately 2 seconds as an average. The best solution of EGAL was a candidate (0.9586, 0.9100). EGAL had approximately 3 seconds as an average.

The running time required to solve this problem is higher than the HS algorithm's running time: obviously EGAL converges more slowly because HS can focus on finding a single optimum, thus it converges faster.

Rosenbrock function:

$$f(x) = \sum_{i=1}^{d-1} \left[100\left(x_{i+1} - x_i^2\right)^2 + (x_i - 1)^2\right]$$

The second Goldstein and Price I. function has four local optima, and one global minimum (0.0, -1.0), and the optimum function value is 3.0.

Goldstein and Price I. function

$f(x)=[1+(x_1+x_2+1)^2*(19-14x_1+3x_1^2+-14x_2+6x_1x_2+3x_2^2]*[30+(2x_1-3x_2)^2*(32x_1+12x_1^2+48x_2-36x_1x_2)+27x_2^2)]$

The number of iterations was set to 20000. We used the following settings according to [2]: *HMCR*=9, *PAR*=0.35, population size: 20, and 20 independent runs. In each tests, HS algorithm found a candidate around the optimum (the average best was: (0.0001, -0.9998), the function value was: 3.0000, approximately 2 seconds as an average. The best solution of EGAL was a candidate (0.0178, -0.9973), the function value was: 3.06523. EGAL had approximately 3 seconds as an average.

In this case the running time required for solving this problem is also higher than the HS algorithm's running time: obviously EGAL converges more slowly because HS can focus on finding a single optimum. And the result are more inaccurate, because EGAL focuses on the diverse in the same time.

Summarizing the former and the current test results: EGAL is effectively focusing on finding a diverse set of solutions, but if we would like to find a single optimum, the algorithm is a little bit slower and less accurate than the original HS algorithm.

Conclusions

In this paper an improved EGAL algorithm was proposed. We showed that this algorithm solves effectively some complicated exercise generation problems. Improved EGAL can generate very different and good quality exercises to test the students' knowledge for different subjects. The behaviors of these kind of problems are summarized in the paper.

The original EGAL can apply for problems with the following behaviors. (1) We would like to find more subsets of predefined questions or conditions or functions. (2) If we check any two possible solutions, we can decide clearly in according to the fitness value, which shows us the better solution. (3)There are too many possible solutions, we can not consider all of them, but fortunately we do not want to find an exact optimum, we are satisfied with a good enough set of solutions if those solutions are enough close to the optimum. (4) We would like to give different tasks for the different students, according to the suitable diversity measure, which was integrated into the fitness function.

The algorithm in this paper was improved by two new criteria (difficulty and group number), they were integrated into the algorithm. Improved EGAL can be applied for problems in the following cases: (a) we have questions or functions or conditions etc. with different difficulty, we would like to ask equally difficult question sets from predefined questions, and/or (b) we have questions from more different groups, and we would like to ask questions from more groups, but not definitely always from every groups. Note that to find the good questions, functions, conditions, the necessary parameter values, difficulty values, groups it is essential to have a good understanding of the given subject, so the teacher and the teacher' knowledge is the "most important parameter" for the algorithm.

At last some test results was given for some widely accepted benchmark functions to examine the difference between harmony search algorithm and EGAL algorithm. We found that EGAL is effectively focusing on finding a diverse set of solutions, but if we would like to find a single optimum, the algorithm is a little bit slower and less accurate than the original HS algorithm.

References

[1] Blanka Láng, Zsolt T. Kardkovács. (2016) Solving exercise generation problems by diversity oriented meta-heuristics, In: Shuang Cang, Yan Wang, SKIMA: 2016 10th International Conference on Software, Knowledge, Information Management & Applications: Chengdu University of Information Technology, China, December 15-17, 2016. Konferencia helye, ideje: Chengdu, Kína, 2016.12.15-2016.12.17. Piscataway (NJ): IEEE, 2016. pp. 49-54. (ISBN:978-1-5090-3298-3) .

[2] K. S. Lee and Z. W. Geem, "A new meta-heuristic algorithm for continous engineering optimization: Harmony search theory and practice," Computer Methods in Applied Mechanics and Engineering, vol. 194,pp. 3902–3933, 2004.

[3] L.C.W. Dixon, G.P. Szego, Towards Global Optimization, North Holland, Amsterdam, 1975.
[4] Ke-Lin Du, M. N. S. Swamy, Search and Optimization by Metaheuristics: Techniques and Algorithms Inspired by Nature, Birkhäuser, 2016
[5] H.H. Rosenbrock, An automatic method for finding the greatest or least value of a function, Comput. J. 3 (3) (1960) 175–184.
[6] A.A. Goldstein, J.F. Price, On descent from local minima, Math. Comput. 25 (1971) 569–574.

Digital Archive System for Utilization of Earthquake-Related Material

[1]Hiroki Tomizawa, [2]Akihiro Abe
[1]Iwate Prefectural University, Japan [2]Iwate Prefectural University, Japan
eMail: tomizawa@iwate-pu.ac.jp, abe@iwate-pu.ac.jp

Abstract

In the aftermath of the Great East Japan Earthquake, the Japanese government recognized an urgent need to construct a national archive system to store information related to the earthquake to ensure effective analysis and handover of such information. Therefore, some regional libraries are archiving the materials using Online Public Access Catalog (OPAC). Such archives contain valuable information; however, it is difficult for most people to source valuable information from an archive as a majority of the material comprises documents with no background, such as white papers, planning papers, handbills, and free papers. In this study, we propose a new digital archive concept aimed for efficient utilization of digital materials adopting the Soft Systems Methodology framework. It is a new concept that connects primary material to new secondary material. The new concept was developed in collaboration with the Iwate Prefectural Library. And we developed a prototype system to verify the feasibility of the framework. We verified that the system has suitable functionality to promote the utilization of the materials, which assisted in solving the problem of the OPAC.

Introduction

The Great East Japan Earthquake and Tsunami, which occurred on March 11, 2011, was a powerful earthquake that caused devastating damages in the Pacific coast of Tohoku region and surrounding areas in Japan. To this day, Japan is recovering from the devastating event. Currently, regional libraries have started archiving mainly paper-based earthquake-related material in an effort to hand over the memories of the disaster to the future generations before they fade away [1]. However, the archives face issues regarding the collection, storage, and utilization of individual documents. Concretely, it is challenging to find the target archive material via the bibliographic records registered in the Online Public Access Catalog (OPAC). Moreover, in a majority of cases, it is difficult to assess the content of documents using solely their bibliographic records. As a vast collection of archive material is already registered in the OPAC, a quick solution for the utility of this collection is unlikely.

We discussed that solving these issues through construction of a new information system would improve the current situation. On the other hands, in each of other researches, the objective is to increase usefulness of the system to particular individuals or groups and they do not consider how to return the relate information of archive materials to an indeterminate number of users.

In this research, we designed a utilization model that revitalized the usage of archived earthquake-related materials as well as that of the archive materials that are derived from them. Using the material of the Iwate Prefectural Library as our subject, we designed the information systems in collaboration with the librarian for improving the utility of

archive material. Moreover, we adopted the Soft Systems Methodology (SSM) that allowed the team to form a worldview and investigate the direction and the course of the project [2]. The model was necessary because detailed information regarding the project's issues and tasks was not shared among the stakeholders; moreover, the objectives towards achieving the practical realization were ambiguous. Our goals are to clarify the process for the model construction as well as develop and evaluate a prototype system to assess the feasibility of the utilization model.

Theoretical background

On October 2012, Japan's Ministry of Internal Affairs and Communications implemented the Great East Japan Earthquake Archive Foundation Establishment Project that served as a platform for accumulating the disaster's records [3]. The National Diet Library's Great East Japan Earthquake Archive, nicknamed "Hinagiku," and OPACs from each library are linked, enabling the cross-searching of archive materials from all of these sources [4]. Although the collection of archive material has been undertaken separately in each region and stored in *Hinagiku*, centralized searching of independently collected data has been largely neglected. Regarding the collection of a variety of earthquake archive materials, Watanabe [5], who is associated with Great Hanshin-Awaji Earthquake Disaster Materials Collection, said "Often, a part of a document is useful, rather than the whole book or a magazine as published, hence the number of archive materials stored in formats such as extracts or clippings have increased greatly," and pointed out that it is difficult to categorize the materials. This statement identifies the difference between libraries that categorize the archive materials according to their contents and archives and museums that place importance on the format, location, and condition. For example, subjects sorted by Nippon Decimal Classification [6] are mainly books; however, most of the archive materials contain different forms of information, and it is difficult to explain the contents of such archive materials using solely their bibliographic records.

Issues facing Iwate Prefectural Library

As the Iwate Prefectural Library (Morioka City, Iwate Prefecture, Japan) is in the disaster-stricken region, it has voluntarily started a collection of materials and initiated a collaborative campaign named "submit your earthquake records to libraries" [7]. On October 2011, earthquake-related materials section in the library was opened to the public (Figure 1).

Figure 1: The earthquake related materials section

As the archive comprised paper-based media resources that required similar management methods as those required for books, it was possible to set up the archive without major changes in the library's operations. The library management has since been taking efforts to increase the number of archive materials and has affirmed its commitment to maintain the archive with paper-based media. In 2011, the year the archive was launched, there were 2833 items. By the end of 2017, the items in archive had increased to 28741, which was an increase by approximately 10.15 times. Books and single-page materials (leaflets, posters, etc.) comprised 55.15% of the entire materials. This trend is expected to continue in the future.

A web portal was also developed to display special exhibitions at the library to prevent the fading of the disaster memories. Great East Japan Earthquake information portal at the library assists in searching for the archive materials in 14 different topics. Single-page archive materials, photographs, and reconstruction plans are all listed as independent topics. Although the OPAC at the library is linked with *Hinagiku* on January 2014, similar to the linking of other projects, a requirement of strategy for the archive material utilization was noted to be essential.

Rerated works

A rise in the affordability of personal computers (PCs), smartphones, and tablet PCs has facilitated an increase in the research process on the effects of information technology on Internet usage. However, no notable development has undergone in OPAC ever since it was launched. Recently, some parties are taking efforts to improve the search functionality user interface and site navigability of social tagging systems [8], using the "Discovery Interface" known as "The Next Generation OPAC" that makes use of the Application Programming Interface (API). For example, Nevzat et al. [9] propose linking of the Social Network Services with OPAC.

Nagai [10], from Tohoku University Library, mentioned that as tasks for utilization of archive materials "make material usage a routine" and "make the work duties a routine." The former quote raises points such as not treating the archive materials as special collections but making them searchable alongside ordinary materials, assisting in using the materials for earthquake education and active learning. The latter acknowledged that there is a need to collect as many archive materials as possible. The Disasters Digital Archive (JDA) of the Reischauer Institute of Japanese Studies is an earthquake-related archiving system that is still being updated [11]. JDA provides a service "Waku Editor" that allows private collections to be created from the archives and their public or private access status to be managed. However, the system does not particularly focus on OPAC-registered materials.

Research process

In this section, we outline the design process for the system that was developed for the earthquake-related materials in the Iwate Prefectural Library. Massive quantities of archive materials have already been registered in the OPAC; thus, it is thought that the development of a new ICT infrastructure can improve the utility of this material.

Therefore, two researchers and seven librarians from the Iwate Prefectural Library were asked to design the new system to facilitate the utilization of the archive materials. We adopted the SSM seven-stage framework [2] that allowed the team to form a worldview and to investigate the direction and the course of the project. This was necessary because information about the tasks at hand was not shared among the team and the objectives towards achieving the practical realization of the project were ambiguous. Thus, SSM seven-stage was implemented as follows:

1. finding out (rich picture)
2. expression the problematic situation (related system)
3. deriving root definitions of relevant systems (CATWOE framework)
4. deriving conceptual models (conceptual model)
5. comparing conceptual models with the real world (DFDs)
6. analyzing feasible and desirable change (materials utilization model) and
7. taking action (prototype system development and evaluation).

We clarified the process in each of the SSM's seven stages based on the details discussed in the eight review meetings that were conducted between 2013 and 2014, those discussed in the email communications (113 emails in total), and the results of each material. In stages six and seven, we evaluated and implemented the prototype system between 2015 and 2016.

Stages 1-2

Finding out the problematic situation

As we discussed**,** the librarians shared information about the OPAC and debated over the OPAC issues. Moreover, we studied the current situation surrounding the archive materials and drew numerous rich pictures regarding the issue at hand (Figure 2). As a result, it was acknowledged that utilizing the OPAC to define the archive materials should be a prerequisite. To achieve this objective, the basic strategy was to design a digital library linked with the OPAC to support the use of the archive materials, which assisted in supplementing the OPAC bibliographic records and in facilitating easy access by many users.

After that, each member of the study group suggested their respective ideas for the new system. This is summarized in Figure 3. From these suggestions, we categorized each item as either a machine process or a human activity. This was done to balance the machine processes and a human activities for the operational direction of the new system. From the categorization, it became clear that the new system would be an information system with a slight bias towards human activities.

Stages 3-4

Deriving root definitions of relevant systems and Conceptual Model

The next step is to perform a CATWOE Analysis for root definition. CATWOE Analysis shows that as long as a sufficient study is conducted, it is possible to create a model, which becomes a fundamental part of the new system:

C (Customers):
Users interested in earthquake related materials
A (Actors): Librarian, researchers
T (Transformation Process):
View archive material→Understand archive material
W (World View):
With background information it should be possible to understand archive materials
O (Owner):
Library（the researcher for the Prototype System）
E (Environmental Constraints):
copyrights on non-book materials etc

Based on the illustration in Figure 3, we attempted to construct a conceptual model that represented the series of activities. When constructing the conceptual model, we took a decision to preserve as much content from the related systems as possible to reflect the wishes of the librarian. As a result, the conceptual model was created successfully, and it was concluded that the new system would include all the features suggested by the study group (see Figure 5). In light of this, as a result of the CATWOE framework as well as studies into root definitions. We defined the new system after the XYZ analysis. That is, the new system is a digital archiving system with OPAC using new materials associated to existing archive materials according to user's theme for activate use of materials.

Stage 5: Comparing models vs real world

As there is a uniform consensus among the study group over the study contents, we proceeded to develop a system analysis using the data flow diagram (DFD) to compare the model with the existing system [12]. DFD is a diagrammatic notation that focuses on exchanges of data between mechanical processes and human activities. Hence, it was thought that finding the true nature of human activities for the new system and utilization of archive materials was required.

Thus, the study group modeled the existing system and identified the functions and roles of the entire system from the flow of data (Figure 6). We then defined the newly provided materials as "new materials," which were distinguished from the existing archive materials. Additionally, we included new archive materials not acceptable in the Earthquake Related Materials Section to increase the utilization of the facility. For example, it is not difficult to imagine that reports, slides, leaflets, and posters created by learners would become great references for new learners focusing on disaster education and workshops for the public or university classes. Based on this assumption, we classified and modeled the users of the new system as either "users" who would view and refer to the new and existing archive materials, or "participating users" who contributed to the promotion of the archive material utilization.

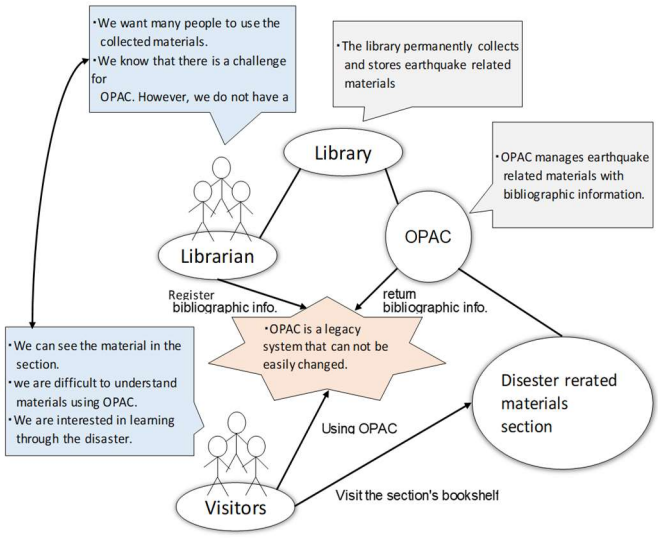

Figure 2: Rich Picture (ex.) .

What kind of system is the Earthquake Archives?

【Human Activities】

- A system to support activities using archived materials.
- A system that can handle copyrights and right to be photographed of the victims.
- A system to promote submission of archive material at risk of being lost.
- A system where anyone can discuss and relive the earthquake based on information accumulated according to their levels of interest and opinions.
- A system to identify the people who could find meanings and utilize the accumulated archive materials.
- A system for everyone to use to consider our relationship with the nature and how to plan for a better future through accumulation of the past, understanding of the present and predict ion and preparation of the future.
- A system that can handle copyrights and right to be photographed

【Machine Processes】

- A system to provide contents according to the user's level of understanding and interest.
- A system to enable recording and viewing of results of activities using archived materials.
- A system to allow registering of additional information related to the archived materials.
- A system to allow advertisements and leaflets created by staffs in charge of Earthquake related materials to be related to archived materials and registered.
- A system that can handle events such as workshops in libraries and outdoors

Figure 3: Problem Expression According to Librarians.
(Relevant Systems)

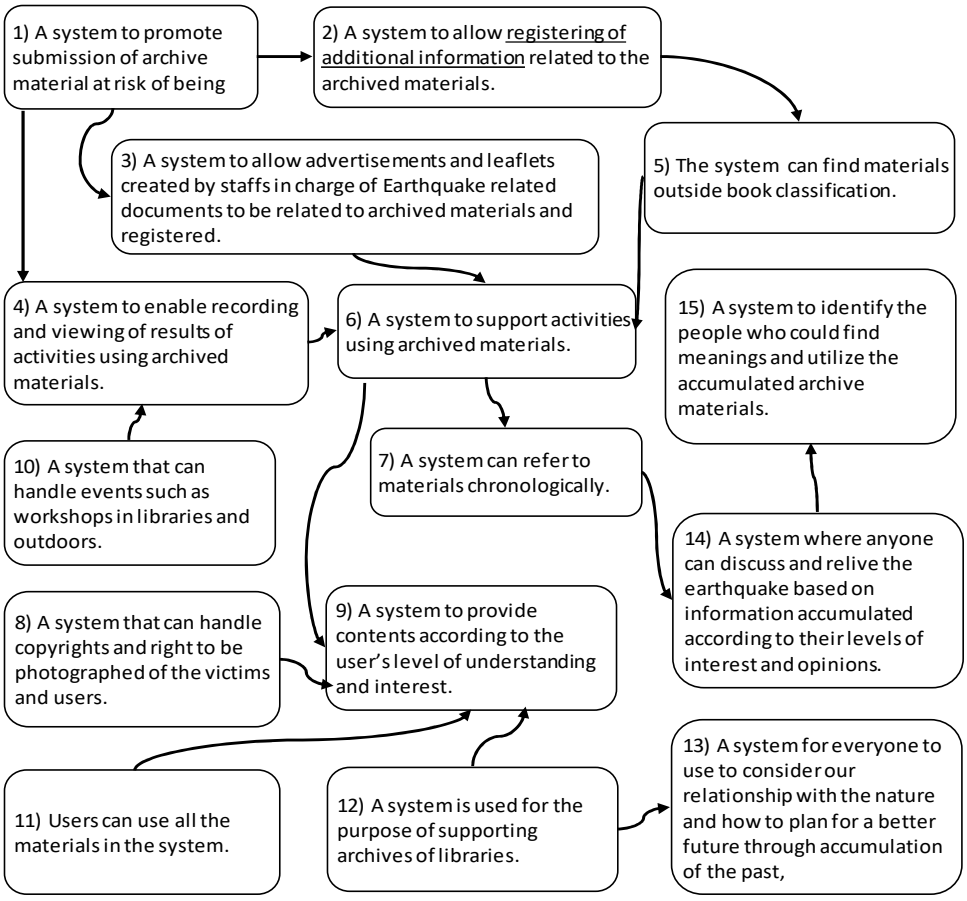

Figure 5: Concept Model.

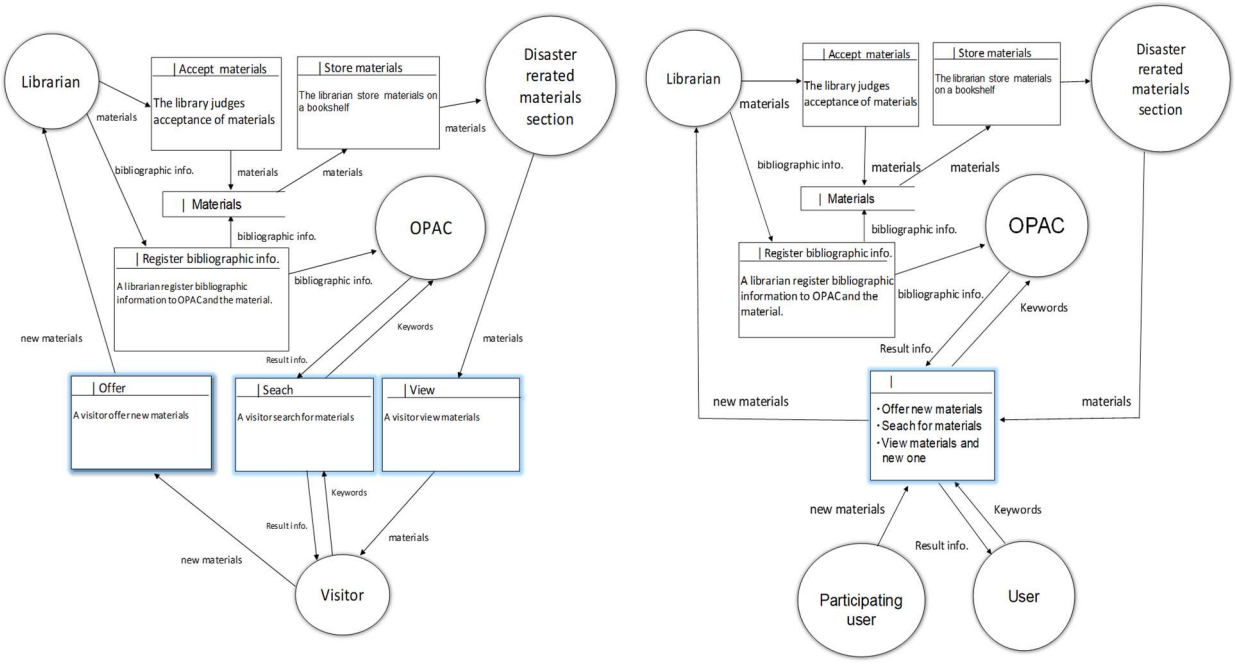

Figure 6: Current Logical DFD (Left) and Requirement Logical DFD (Right).

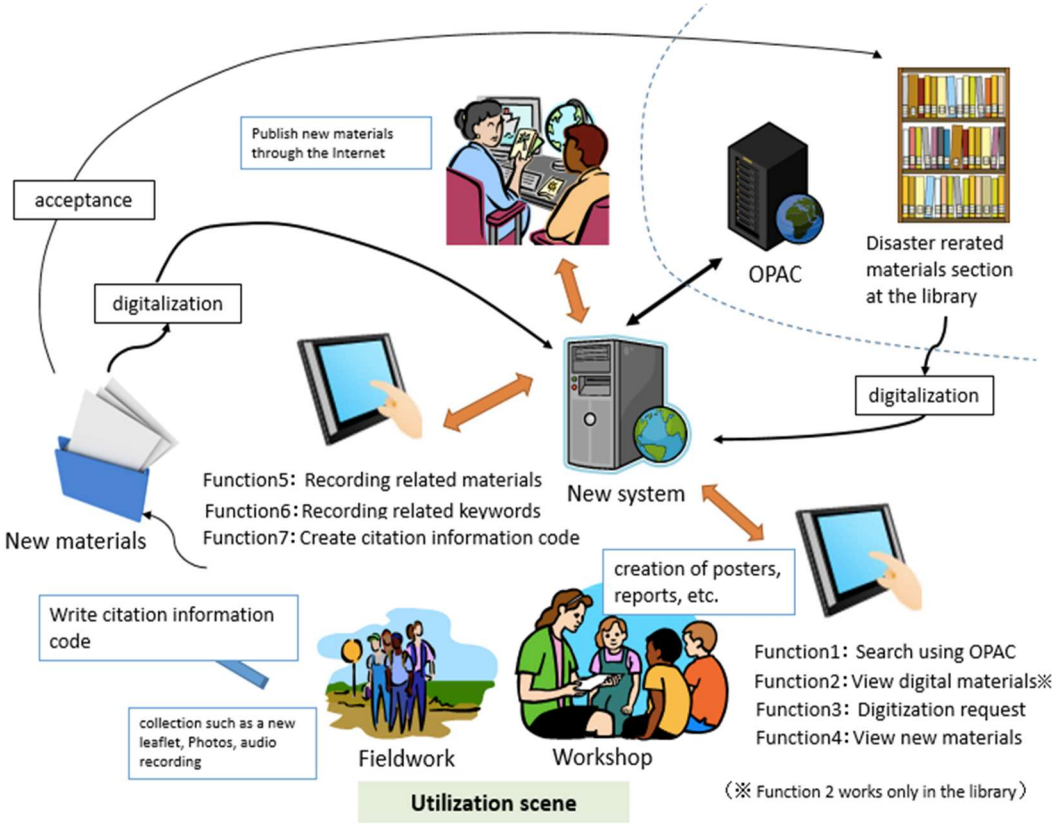

Figure 7 : Archive Material Utilization Model

Stage 6: Analyzing changes

In this study, we defined the reform plan as the utilization model derived from the archive materials (Figure 7). This utilization model is an information system to support the creation of new archive materials, serving as the hypothetical model for this research. In this model, the information in the archive materials was augmented by the new archive materials to develop the machine-supported human activities.

Although the environmental constraints (Figure 4 E) restrict the use of archive materials that have been outside the libraries, the existing archive materials and new materials can be utilized in the digital archive. Moreover, we developed the idea to manage the existing and new archive materials using unique codes named "citation information code" related to a separate reference list. This concept was developed among the reviewed members, which was decided to be "like an electronic medical record." For example, if each new archive material was assigned a unique code, not only would it allow single-page materials or photo data to be easily associated, it would allow a user to find new materials of interest.

Stage 7, action 1: Prototype development

Following the reviews of Stages 1–6, the specification of the new system was decided. We adopted the Cross-Searching System already operated by the Iwate Prefectural Library as the Web API. This core functionality is used to create new digitalized archive materials developed by New Archive Material Providers based on the existing archive materials. The system was installed on the Iwate Prefectural University servers (OS: FreeBSD, Server: Apache, MySQL, Development Language: PHP5). A visit to the Earthquake-Related Materials Section at the Iwate Prefectural Library, using tablet devices, shows that it resembles an end user platform. Thus, this was developed as a web application to be compatible with the desired platforms. The functionalities developed are aggregated into three categories: "Search for Archive Materials," "Create New Archive Materials," and "Learn about Archive Materials" (Figure 8).

"Search for Archive Materials" exchange parameters with the Cross-Searching System at the Iwate Prefectural Library and allows searching for the document title, author, publisher, and year of publication. These also record the keywords entered by the user. The search result page allows viewing the archive materials digitally, which are marked by icons.

"Create New Archive Materials" displays a list of clipped archive materials in the order that has been added, and it provides radio buttons to indicate the importance of each clipped material to the new archive material. The icons on the right displays "Request" button for those materials not yet digitized, and librarians can refer to the requests when processing user's rights for the materials. Clicking a material already digitized takes the user to the viewing screen.

Finally, "Learn about Archive Materials" displays a summary (Creator Name, Nickname, Group Name, Age, Comments, and Thumbnail Image) of new materials in the order referenced. When an item of interest is clicked, the contents of the new archive material and OPAC archive materials, referenced during its creation, are listed in the order of importance. In this list, the same icons, as used in the "Search for Archive Materials" menu, allow users to click the archive materials of interest from this list. The photographic data for each spot is linked to the new archive materials in the system. Photographic data taken with fieldwork may be archived in the library to form new materials.

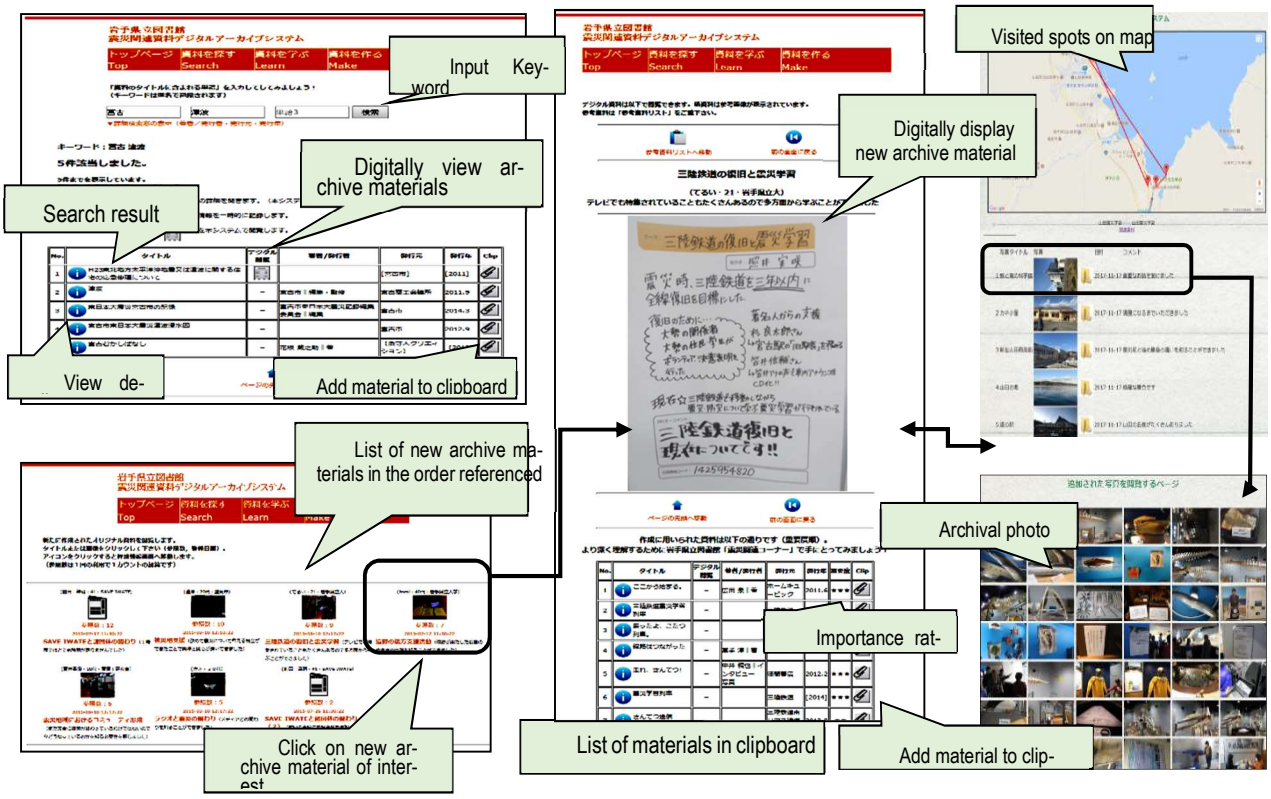

Figure 8: New system example screenshots

Stage 7, action 2: Evaluation

A new program implementation was developed with an additional step of conducting local research (Miyako City, Iwate, Japan). The participants take part in and validate the experiments (total of 4 rounds) that were conducted between November 2015 and February 2016 (Figure 9). They were recruited via word of mouth, leaflets (placed in public facilities), and mass media, among others. The number of participants who responded to the adverts and students interested in learning the Earthquake was 14. The participants took part in the four rounds of workshops with three hours for each round (one day for local research). The sessions were conducted on 25th November (explanation of the project and trialing of the Prototype System), December 12th (local research in Miyako City), January 16th (review and creation of new archive materials), and February 7th (Presentation).

Figure 9: Scenes of workshops and fieldwork.

In the fourth round of the implementation program (February 7th), seven of the participants (two of whom did not participate in the local research) used the Prototype System to put together and presented new archive materials. The chosen topics were “Tide Embankments and Sluice Gates,” “State of Rural Areas, at the Time of the Disaster and Present,” “Lost Town – Model Reconstruction Project,” “Miyako Local Specialty Squid Crackers and Disaster Relief,” “How to Avoid Future Tsunami Damages,” “Implementation of HUG (a Refuge Navigation

Game) and its Link to Municipal Refuge Operation Manual," and "Menus of Meals Provided After the Earthquake".

The satisfaction rating from each round was high. The fact that the participants visited the disaster site during the local research made them realize the importance of utilization of the archive material. It also demonstrated that it is possible to run implementation program over multiple days.

However, we received operability and usability feedbacks from the participants on the functionality of the prototype system. Regarding the usability improvements such as smartphone compatibility, we acknowledged there is a room for improvement in the functionality, such as providing more search support. In particular, the registration of the archive materials found on the library shelves was found to be labor intensive and simplification of this process is required in future. Regarding the implementation of the program, the points raised included that it was difficult for the participants to understand the purpose of the program and its contents during the recruitment stage.

Result and Conclusion

In this study, we focused on the earthquake archives based on OPAC constructed in libraries that were rarely discussed in the previous studies. Then, we developed a new system for Earthquake-related documents to promote its utilization.

A hypothetical model for the archive material utilization model (Figure 7) was developed to solve these issues using the SSM process. The Prototype System was created to evaluate the feasibility of the model; then, the workshop program and implementation program were tested. As a result, we evaluated that this model and the Prototype System were largely effective in solving the issues of document utilization.

For future work, the system should be linked with OPAC, and further improvement must be made in the archive material's search functionality. It is necessary to make the system easier for users to find relevant archive materials. Moreover, the program implementation must be improved by a librarian for effective operation. Finally, improvements must be made for the application of the new archive materials. By clearly displaying the relationship between new and existing archive materials, it will possible to extract additional values from archive materials.

Recently in Japan, programs developed by libraries to encourage a user's contribution are increasing [13]. While the fading of the memories of the Earthquake is a concern, the digital archive described in this research provides possibilities of being in close association with local communities.

Acknowledgments

This research was supported by JSPS KAKENHI Grant Number JP16K00439. We gratefully acknowledge the work of librarians in Iwate Prefectural Library and past and present members of our laboratory.

References

[1] Tohoku University Library: Campaign for Sharing Between Libraries "Donate Your Records of the Earthquake to Libraries", Source: http://www.library.tohoku.ac.jp/shinsaikiroku/ (See: 2018-10-01 (in Japanese).

[2] Peter Checkland, Jim Scholes: Soft Systems Methodology in Action, Wiley, 1990.

[3] Ministry of Internal Affairs and Communications: Promotion of "Digital Archives" Usage. "Great East Japan Earthquake Archive" Foundation Establishment Project (web), Source (http://www. soumu.go.jp/menu_seis aku/ictseisaku/ictriyou/02ryutsu02_ 03000092 .html, See: 2018-10-01).

[4] National Diet Library, Japan: Hinagiku - NDL Great East Japan Earthquake Archives, Source: http://kn.ndl.go.jp/, See: 2018-10-01)

[5] Takahiro Watanabe: Creation of Electronic Archives for Kobe University Digital Library System, Digital Library Issue 16 (1999), pp.3-11 (in Japanese).

[6] National Diet Library: Japanese Decimal Classification Method 9th Edition Classification Standards (Web), Source http://www.ndl.go. jp/jp/library/data/pdf/NDCbunruikijun2010.pdf, See:2018-10-01).

[7] Iwate Prefectural Library: Great East Japan Earthquake Information Portal (Web), Source http://www. library. pref.iwate.jp/0311jisin/, See: 2018-10-01) .
[8] Ken Kuboyama: Trends in Next Generation OPAC: Its Functionality and Prospects in Japan, Information Science and Technology, No.58-12 (2008), pp.602-609 (in Japanese).
[9] Nevzat Özel, Tolga Çakmak：Users' Expectations on Restructuring OPACs through Social Network Applications, IEEE-ACM International Conference on Green Computing and Communications and International Conference on Cyber, Physical and Social Computing, (2010) pp.798-803.
[10] Shin Nagai: Campaign for Sharing Between Libraries "Donate Your Records of the Earthquake to Libraries" The Current State and Issues Faced in Great East Japan Earthquake Related Material Collection by Campaigners: How to apply the experience from the disaster, Japan Library Association, Current Awareness, No.319, (2014) pp.11-13. (in Japanese).
[11] Reischauer Institute of Japanese Studies at Harvard University: Japan Disasters Archive (Web), Source: http://jdarchive.org/, See: 2018-10-01) .
[12] Geoff Cutt: Structured Systems Analysis and Design Methodology, Van Nostrand Reinhold Computer, 1988.
[13] Chika Igaya: Connected Libraries – Attempt to Become Cores of Communities, Chikuma Shinsho 2014 (in Japanese).

Relationship of Organizational Culture, Leadership Style and Cloud Computing

[1] VIKTOR SPILÁK, [2] TIBOR ZSOLT KOSZTYÁN
[1]OTP Hungary [2]Pannon University
eMail: viktor.spilak@icloud.com, kzst@gtk.uni-pannon.hu

ABSTRACT

Information technologies and the provided advantages play crucial role in companies' lives and affect to their success. The primary goal of the study is to investigate interrelationship among organizational culture, leadership style and cloud computing. The former researches examined only one or two areas thus they didn't try to understand the real operation. The author had worked out a new model which could help in the future that the companies would be able to understand how new information security questions and cloud technologies can affect the organization culture and leadership style.

Introduction

Companies are seeking how they can be better and more competitive. Information technologies have key role that they are able to reach better performance then another player of the market [1]. These technologies can help them to identify and correspond to requirements of their customers [2]. The research is timely because continuous competition is forcing the company that they have to reach advantages if they would like to survive nowadays. Information Technologies are important ingredients in this process because they are able to ensure chance to speed up the delivery and reach the customer faster than they competitors [3]. Thus technologies have become a specific tool which can increase the ability of an organization to compete better but if they use them inefficiently it will cause the opposite effect. The leaders try to use these technologies cost effectively therefore the cloud computing can be an alternative because it ensures common used resources, more effective operation and flexibility [4].

Researchers had examined organization culture, leadership style and cloud computing as individual areas but they didn't establish a common model where they can analyze how these areas can affect to each other. If we can establish a connection among these areas it can help us to understand how can it affect one information security or cloud computing related question to the whole culture and leadership style of the organization.

The next part of our research we are reviewing the relevant technical literatures and we are showing our new model. Later we are presenting our case study furthermore the reached results of our insight.

Literature review

In this section we demonstrate the most important articles related organization culture, leadership style. We summarize information security related requirement and criteria which is mandatory if an organization would like to use the cloud base services.

Culture models

The organization cultures related researches are a very colorful topic thus we can identify many models [5]. Many different approaches were born during years. The differences among these purposes are which areas were prioritized by researchers [6]. Earlier the organization culture was defined as the most important factor. Later this theory was too excessive but experts agree this item is one of the key indicators in the life of a company [7]. The culture can affect the performance of the organization or it has a positive effect if employee identify oneself with the culture and the goal of the company [8]. The culture is crucial because it forms leaders who can influence which culture will be established inside [9].

Related examined models we have identified which are capable of building the connection among the information security and cloud computing. The base of our establishment were related dimensions or evaluation criteria contain any technical areas which are fit to ensure opportunities to connection. As per this approach Handy culture model, Cameron-Quinn Competing Value Framework and Henry Mintzberg organization culture are suitable.

The leader of a company has to create and provides proper working environment, develops abilities of employees, and as the culture creator supports his/her staff to reach their own goals [22]. These management dimensions include functions of the leader, problem solving process, leadership style, control systems and role of leaders as well [23]. Because the leadership style has outstanding effect to form an organization therefore we assume it has significant influence to the examined areas by us too. For this reason we gave special attention to involve it in our research.

We have chosen the Cameron-Quinn Competing Value Framework because it is appropriate to analyze not only the organization culture but it can define the leadership style as well. Further advantages of this method are functional and we can apply it during our research without we get too deep insight into the organization life. It could be negative because increasing complexity of our work will distract from key questions. The framework ensures different approaches of evaluation and creates opportunity to measure current and desired culture as well. It is essential because we can analyze the strategy thinking as well in later stage of our investigation. The model is capable of identifying strengthen of the cultures and relations with characteristic of organization. The Framework which was established by Cameron and Quinn can resolve the culture and leadership style simultaneously thus it is able to decrease the complexity of our work.

The OCAI (Organizational Culture Assessment Instrument) questionnaires was established by Cameron and Quinn and help us to classify the organization by four culture types and six dimensions. These include the dominant characteristic of the organization, the typical leadership style, the cohesive force, the strategy focusses and the success criteria [29]. The OCAI questionnaire beyond that identify the dominant culture, can show the harmony within the organization [6].

Table 2: Culture models

Culture models	Possible connection with technical areas	Investigated areas/company values/ main focus
Handy's culture model [10]	**Suitable** Take into account the applied technologies	The model focus on past of the organization, ownership, goals, applied technologies, environment and people.
Hofstede - National culture model [11]	**Not suitable** It investigates only the relationship among people	The power distance and the individualism/collectivism define the activity pattern.
Morgan's culture model [12]	**Not suitable** The technology aspect is not part of this structure.	Leaders think the organization is not more than a very specific machine or living body, which is adapting to the environment during its live.
Trompenaars' culture dimensions [13]	**Not suitable** The culture types were generated by structure of organization (vertical and horizontal nature) and the connection among employees	The culture types are between two horizontals and two verticals. And these dimensions that a type is personal or task oriented perhaps hierarchical or egalitarian.
Cameron – Quinn's Competing Values Framework [14]	**Suitable** The model handles the technology as a success factor.	The behavior of organization is defined by level of stability, flexibility, control, dynamism, self-determination, internal or external focus. The value preferences are a key indicator.
Wong's model [15]	**Not suitable** Applied dimensions are human behavior, relationship with each other.	The model distinguishes doer and being oriented cultures. This approach focuses on nature, time, activity, communication, space, power, individualism, competitive spirit, organization, formality.
Hall's model [16]	**Not suitable** Technology only appears as a transferring tool.	Interpretation of information depends on common knowledge. Value and mentality of the culture are part of this thinking. Different ways of communication are important.
Slevin – Covin's model [17]	**Not suitable** ICT has only a small part of the whole model.	The model shows how the organization can adapt to the continuously changing environment. It distinguishes organic and mechanic culture.
Harrison's culture model [18]	**Not suitable** Identified areas don't cover the technologies as main characters of an organization.	It identifies four culture types where difference is defined by connection, point of applied structure, leadership and control. The organization specialties are created by control, motivation, central values and negative consequences.
Henry Mintzberg's organizational configurations [19]	**Suitable** The structure contains all employees who are responsible for the IT and financial systems.	The model defines five areas which specify the value and operation of an organization. These are core of operation, strategy level, technostructure as a middle and support staff.
Kluckhohn and Strodtbeck's culture dimensions [20]	**Not suitable** It focuses only on the human activities therefore it is not possible to build a connection with another disciplines.	It investigates six attributes: personality, relationship among people – nature and people – people, activity of the community, space and time as well.
Globe cultural dimensions [21]	**Not suitable** It doesn't calculate with technologies and their influence within the organization	It examines how cultural values are related to organizational practices, conceptions of leadership, the economic competitiveness of societies, and the human condition of its members.

Table 3: Leadership style

Models of leadership style	Investigated dimensions/evaluation criteria/main focus
Henry Mintzberg [19]	This model is searching which role is covered by leaders. It depends on the leader's place in organization hierarchy or where he/she is working. Three leaders' roles are identified as interpersonal, informal and decision maker.
John Kotter [24]	Manager – who does thinks well, Leader – who does good thinks It handles the leadership and management as an independent notion.
Dian Hosking [25]	The features of management are rational evaluation of situation, systematic development and aligning necessary resources. The generalized roles of the leadership ensure circulation of creation, production, re-production, conversion activities.
Abraham Zaleznik [26]	It distinguishes manager and leader roles. Manager minimalizes numbers of options but the leader is searching continuously new opportunities and approaches. Manager accepts goals of organization but the leader creates them. Manager is less emotional than leader.
Richard Boyatzis [27]	He separates three different dimensions: group of competencies, functions of management and organizational environment. Competencies are defining characteristic of a person.
Quinn [28]	He is trying to define the orientation of leadership mentality. It could be internal or external (environment) focused. The model stated that it is not either-or type characterization but leaders have different nature in the same time.

Cloud computing-based systems

Cloud technologies promise a completely new way to deliver the IT services [30]. Companies can apply information technology on demand which can deliver better flexibility and cost-effective usage [31]. This new approach is demanding a fundamental new thinking and operation model furthermore it is capable of ensuring a higher availability or decreasing the costs [32]. During our study we have been analyzed that models which are focused not only technologies and technical aspect but also tried to understand the effect to the information technology and the IT himself. Part of our analysis we have taken into account five factors as cloud approach, operation, service model, information security management as well as main evaluation criteria.

In the center of processed models are the level of digitalization, the information technology and capability of the cloud. The maturity level of particular organization is defined by the combination of maturity criteria. We have identified the main problems are not in technical solutions but they are connected to the management and the leadership style. Which models were appeared after 2017 contains management related approach as well. Because of there are capable of helping the leadership team to make decision and support to integrate the cloud base services into the daily workflow. It is important to highlight that these models aren't be able to define in every case how companies have to apply cloud services because it depends on the characteristic of the organization. As per investigated researches we defined our own model which consist of cloud model, operation, services and information security management.

Earlier the culture model, leadership style and cloud computing were not handled in the common framework thus we don't have exact information about the relationship between them. We have found some attempt that one research areas would be connected to another but they have focused only some smaller areas. Our purpose was that we can establish a common model which helps for the companies to understand the relationship between these extremely important disciplines. Our research model (Figure 1) is containing not only the cloud area but we had involved the information security as well because it will be an important question on the later stage of our research. But into this article we are focusing only the cloud related questions.

Case study

We would have liked to investigate one of Hungarian Telecommunication companies. These providers cannot ensure their services without information technologies therefore we are absolutely confident

that they try to implement parallel the stable and state-of-the-art technologies as well. Our goal was that we can explore how the cloud base services can affect to the organization (with particular attention related IT and information security management).

Research questions

- **Q1:** Which kind of requirements are defined against the cloud services?
- **Q2:** How will cloud services affect to information security management?
- **Q3:** Which kind of changes were driven by cloud services into organization culture and leadership style?

Table 4: Cloud computing related models

Maturity model	Applying cloud based solutions				Examined dimensions/main assessment criteria
	Cloud model	Operation	Services	Information security manage-	
Oracle: Cloud Computing Maturity Model Guiding Success with Cloud Capabilities [33]	x	x	x	x	It determines maturity of organization by strategy, architecture, infrastructure, information, operation-management, project portfolio as well as evaluation of organization and guidance.
The Road to The Responsible Cloud [34]	x	x	x		On the basis of periodical improvement and management it evaluates each maturity levels.
Managing Cloud Computing: A Life Cycle Approach [35]	x	x		x	By means of Lifecycle management, it would like to control not only the introduction of cloud services but daily operation of public cloud.
Towards a Consumer Cloud Computing Maturity Model - Proposition of Development Guidelines, Maturity Domains and Maturity Levels [36]	x	x		x	It shares domains of cloud models to organizational and technical groups. It also touches upon the regulation, the safety, the organizational readiness, processes, infrastructure, and management of operation.
Cloud Maturity Model [37]		x	x		Outsourcing lifecycle and CMMI (Capability Maturity Model Integration) creates the fundament of cloud maturity model.
Cloud Computing With a Model Futuristic Maturity [38]	x	x	x		It suggests phase approach as well as defines five key components: consolidation, virtualization, automation, use and cloud.
Cloud Data Governance Maturity Model [39]	x	x	x	x	It focuses on the examination of data handling in connection with cloud solutions.
Enterprise Cloud Adoption - Cloud Maturity Assessment Model [40]		x		x	It identifies 11 key components which have influence on the introduction and using of cloud solutions.
Maturity Level of Cloud Computing at HCT [41]	x	x			It applies service guided architecture approach in measuring the maturity model.
FHNW Maturity Models for Cloud and Enterprise IT [42]	x	x	x	x	It doesn't follow the usual evaluation of cloud maturity model. It focuses on why they use cloud solutions and by what means it changes tasks and positions of IT.

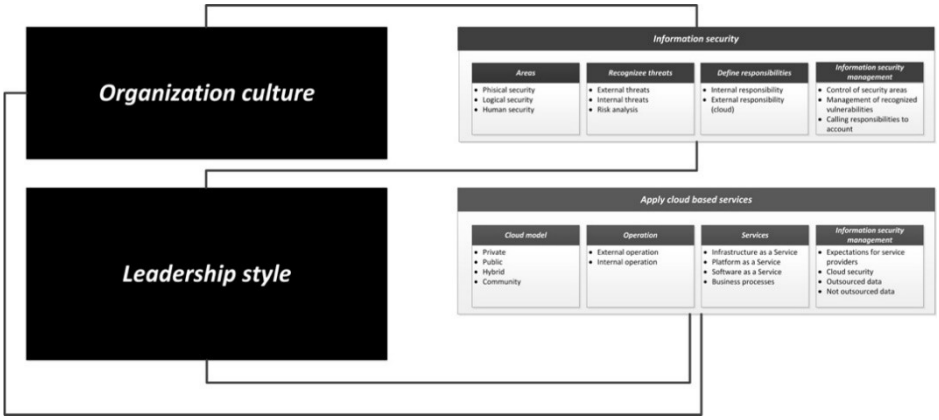

Figure 1: Research model

Table 5: Sampling characteristic (N=6)

ID	Position	Key responsibility	Team size
RE1	Application manager	Ensure the continuous operation for all applications with internal and external resources, provide proactive and reactive troubleshooting.	10-20 employees and vendors
RE2	Infrastructure operation and development manager	He is responsible to operate and develop network, firewall, server, storage, database, middleware, backup systems.	20-25 employees and vendors
RE3	Application development manager	Prioritizing and executing the development needs of business areas and application management.	25-30 employees and vendors
RE4	Desktop support and helpdesk manager	He is responsible to manage the L1 teams and dispatch income ticket to the proper L2-L3 experts. The team maintains all desktops and notebooks.	15-20 employees
RE5	Governance leader	Coordination of finance and managing regulatory and audit tasks.	5 employees
RE6	Information security leader	Executing the vulnerability tests, managing revealed deficiencies with application and infrastructure teams. They investigate and evaluate appearing security risks.	2 employees

Sampling in the qualitative section

In the data collection we have been using the qualitative method because our purpose was that we could understand deeply the thinking of the IT organization. We executed query of structural questions with all IT leaders in the interest of get the exact picture of the organization. The examination units were different areas of the IT.

Qualitative data analysis

As we found into the literature review cloud computing has different advantages as good scalability, flexibility, automation and high availability which were confirmed as per our qualitative research. But our responders didn't highlight the cost-effectiveness thus we conclude that the scalability and the availability are the most important to these kinds of companies.

> "Easy operation, scalability and flexibility are extremely important in our market. The cloud base solutions help us to establish a homogeneous operation." (RE2)
>
> "Distributed structure and central management could ensure a better stability for these services." (RE1)

These services have great deal of advantages but it was important the cloud service model (private, public, hybrid or community), type of service (IaaS, PaaS, SaaS, etc.), method of provider selection and the operation approach (internal or external).

> "We can find different cloud services and approach on the market. We can buy public services and build an own private cloud as well. As per our opinion the hybrid cloud could be the future of IT." (RE2)
>
> "Thus, we can keep our critical data inhouse and we can get the scalable, flexible and limitless resources from cloud provider in the same time." (RE6)

The responsibility of information management organization was changed when the company started to take into consideration cloud base solutions. Earlier this team focused only to internal applications and services. But now they had to implement a completely different mindset because they should ensure rules and continuous control over the cloud services that data will be in safe. But they cannot affect to big vendors in every case. Therefore, the internal function was changed and they try to become a consultant who can suggest the secure solution for the business but the earlier function was not completely disappeared.

> "Before cloud services the role of information management was to protect our internal solutions against hackers and sort out every possible vulnerability." (RE6)
>
> "New requirements, that generated by management and operations, forced us to change. We had to understand that private cloud has lots of advantages which able to deliver improvement for our company too. Ergo we cannot say that we don't support these solutions." (RE5)
>
> "If we would like to support the organization we need to change our mindset and functions as well. For this reason, we have involved new competencies into our team. We have to help our internal client to define which are the safe and approved cloud services and which are not. But it is not enough because the most important is to define the data set which are cloud ready or which are stored compulsory in-house." (RE6)

These additional functions justified to hire new members on information security area. But the management team couldn't accept this fact in beginning but later they had to realize the necessity of this step.

> "We could extend our team only after internal users started to use cloud services without IT approval. Thus, lots of business-critical data were leaked therefore the management decided to strengthen our team." (RE6)

The public cloud services have to be compliance with internal security requirements before the company decides to use them. But these security criteria were worked out related only internal applications therefore it was necessary to build one for external services. It should contain how, when and where the organization able to use the public cloud.

> "We defined three different stages of cloud services that we can assure distinct security level for different classified data (Private cloud in Hungary, International Private Cloud and International Hybrid Cloud solution)." (RE5, RE6)

Cloud base services have clear affect not only on the information management team but they are able to change the tasks of operation members. If a company own an onsite infrastructure, they need to have a group who are able to manage these infrastructures physically as well. The situation is same if we consider the data center as well. But the cloud base services (IaaS, PaaS, SaaS) don't require these tasks thus it could impact negatively to whole IT organization. This effect depends on the service level of the cloud. Because IaaS could eliminate the data center related tasks, the PaaS is connection with the infrastructure related task until the SaaS able to affect negatively the whole IT stack.

> "We have to admit the public cloud can touch us and completely redefine our operation model. Lots of activities will become unnecessary because the cloud provider will do it itself." (RE1)

> "I don't think so the internal IT will be completely terminated because we have lots of requirements if we would like to use public cloud. I believe in that hybrid cloud will be the future where the internal IT will have a determinative role. But experts have to adapt to these new circumstances." (RE2)

Some jobs will become more important than earlier. They are existing in the current structure but their focus will be changed (for example: service manager). Our responders assume that the management layer will not be vanished but it's role will be transformed.

> "We handle SaaS services as an internal one and assigned to the appropriate manager, for example IaaS is managed by infrastructure leader. We strive to use the same KPIs as internal solutions." (RE5)

The investigated company has an adhocracy culture as a dominant. It is an advantage because the dynamic working environment could be success factor in a telecommunication sector. As per our experience employees take the risk that they can reach an outstanding result and the innovation is extremely important them. It could be a good environment to apply cloud base solutions but as per this result we cannot say that this culture has significant benefit to apply the new technologies.

Results

Our investigation shows that the adhocracy culture is essential that a company can be success on a competitive sector. The result-oriented approach motivates the company to implement new innovative technologies. These solutions are able to help them to decrease costs or improve the efficiency (Q3). Cloud based services can help to focus to core competencies of companies. It could be an important advantage because they can invest to their business and avoid IT costs. Last few years a culture change was realized despite it is a long process. The responsibility of information management team was transformed. They have to be active participant in selection process of the cloud-based solutions and they have to continuously verify the necessary security level of these services (Q2). The organization defined the security requirement against the cloud providers. It was an important step because they can avoid to use an unapproved application by their users (Q1). But it is not an easy process because the IT had to defined proper alternatives to their users before prohibit any public services. The investigated subsidiary has defined standards and building blocks which are mandatory to use when they are building a private cloud (Q1). The IT has to realize an organization challenge if they use cloud related services. Some role will be transformed another will be terminated permanently (Q3). It will not mean that this skill will be terminated on the whole IT market. They will concentrate to the cloud providers and that companies which wouldn't like to use cloud-based services. But experts must adapt to a new situation and this capability will specify their future possibilities on the job market (Q3).

The next step of our research we are going to extend our examination with information security management and verify our results and assumptions on a wider sample. We are going to work out our questioners because the structured interview is not suitable to execute this data collection. The current approach was proper because we would have liked to get deep information and context from one company but it will not be proper methodology when we will investigate thousands.

References

[1] C. Yang , W. Yi, N. Saggi, B.-A. Jose and . K. Gang, "IT capabilities and product innovation performance: The roles of corporate entrepreneurship and competitive intensity," *Information & Management,* pp. 643-657, 2015.

[2] N. Roberts and G. Varun , "Leveraging Information Technology Infrastructure to Facilitate a Firm's Customer Agility and Competitive Activity: An Empirical Investigation," Journal of Management Information Systems, pp. 231-270, 2014.

[3] K. Marchese, J. Crane and C. Haley, "3D opportunity for the supply chain: Additive manufacturing delivers," *Deloitte University Press,* 2015.

[4] P. Fehér, A. Kő and Z. Szabó, "Kapacitásmodellezés és az IT-architektúratervezés kihívásainak vizsgálata statisztikai és prediktív analitikai eszközökkel," Statisztikai szemle, pp. 1149-1164, 2016.
[5] O. Oju, "Impact Assessment of Corporate Culture on Employee Job Performance," Business Intelligence Journal, vol. 2, pp. 388-397, 2009.
[6] Á. Balogh, Z. Gaál and L. Szabó, "Relationship Between Organizational Culture and Cultural Intelligence," *Management & Marketing Challenges for the Knowledge Society,* vol. 6, no. 1, pp. 95-110, 2011.
[7] J. C. Naranjo-Valencia, D. Jimenez-Jimenez and R. Sanz Valle, "Innovation or imitation? The role of organizational culture," *Management Decision,* 2011.
[8] O. Oju, "Organizational Culture and Corporate Performance: Empirical Evidence from Nigeria," Journal of Business Systems, Governance and Ethics, vol. 5, no. 2, pp. 88-100, 2010.
[9] Z. Gaál, L. Szabó and N. Obermayer-Kovács, "„Tudásmenedzsment-profil" érettségi modell," *Vezetéstudomány,* 2009.
[10] C. B. Handy, Understanding Organizations, Harmondsworth: Penguin, 1999.
[11] G. Hofstade, Cultures and Organizations: Software for the Mind, London: McGraw-Hill, 2010.
[12] G. Morgan, Images of Organization, London: SAGE Publications, Inc, 2007.
[13] C. Hampden-Turner and F. Trompenaars, "Cultural Intelligence: Is Such a Capacity Credible?," *Group & Organization Management,* pp. 56-63, 2006.
[14] K. S. Cameron, R. E. Quinn, J. DeGraff and A. V. Thakor, Competing values leadership: Creating value in organizations, Northampton: Edward Elgar Publishing, 2007.
[15] Z. Gaál, "Emberi tőke – szervezeti kultúra," *Harvard Business Manager ,* pp. 69-76, 1999.
[16] P. S. Tolbert and R. H. Hall, Organizations: Structures, Processes and Outcomes, London: Routledge, 2008.
[17] J. G. Covin and D. P. Slevin, "Juggling enterpreneurial and organizational structure. How to act together," *Sloan Management Review,* pp. 43-53, 1990.
[18] G. Caroll and R. J. Harrison, "Organizational demography and culture: Insights from a formal model and simulation," Administrative Science Quarterly, pp. 637-667, 2005.
[19] H. Mintzberg, A menedzsment művészete, Budapest: Alinea Kiadó, 2010.
[20] F. Kluckhohn and F. L. Strodtbeck, Variations in Value Orientations, London: Greenwood Press, 1973.
[21] R. J. House, P. J. Hanges, M. Javida, P. W. Dorfman and V. Gupta , ulture, Leadership, and Organizations: The GLOBE Study of 62 Societies, California: Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies, 2004.
[22] L. Szabó and G. Dancsecz, "A nemzetközi sportrendezvény-szervezési projektek sikertényezői és a siker megítélésenk kritériumai," *Vezetéstudomány,* pp. 30-31, 2009.
[23] M. Dobák and Z. Antal, Vezetés és szervezés, Budapest: Akadémia Kiadó Kft., 2016.
[24] J. P. Kotter, Leading change, Boston: Harvard Business School Press, 2012.
[25] D. M. Hosking, "Organising, leadership and skilful process," *Journal of Management Studies,* pp. 147-166, 1988.
[26] A. Zaleznik, "Managers and Leaders: Are They Different?," *Harvard Business Review,* 1992.
[27] D. Goleman, R. Boyatzis and A. McKee, Primal Leadership: Realizing the Power of Emotional Intelligence, Boston: Harvard Business Review Press, 2003.
[28] R. E. Quinn, D. Bright , S. R. Faerman, M. P. Thompson and M. R. McGrath, Becoming a Master Manager: A Competing Values Approach, New Jersey: WILEY, 2015.
[29] K. S. Cameron and R. E. Quinn, Diagnosing and changing organizational culture: based on the competing values framework, San Francisco: Jossey-Bass, 2011.

[30] N. Sultan, "Cloud computing for education: A new dawn," *International Journal of Information Management,* vol. 30, no. 2, pp. 109-116, 2010.
[31] L. Grace, "Basics About Cloud Computing," Software Engineering Institute Carniege Mellon University, Pittsburgh, 2010.
[32] Educause, "7 things you should know about cloud computing.," Educause, 2009.
[33] S. Mattoon, B. Hensle and J. Baty, "Cloud Computing Maturity Model Guiding Success with Cloud Capabilities," Oracle, California, 2011.
[34] D. Drogseth, "The Road to the Responsible Cloud," EMA - IT & Data Management Research , Industry Analysis & $, Boulder, 2011.
[35] G. Conway and E. Curry, "Managing Cloud Computing: A Life Cycle Approach," in *2nd International Conference on Cloud Computing and Services Science*, Porto, 2012.
[36] D. Weiss, J. Repschlaeger, R. Zarnekow and H. Schroedl, "Towards a Consumer Cloud Computing Maturity Model - Proposition of Development Guidelines, Maturity Domains and Maturity Levels," in *Pacific Asia Conference on Information Systems (PACIS)*, Jeju Island, 2013.
[37] A.´. Duarte and M. Mira da Silva, "Cloud Maturity Model," in *2013 IEEE Sixth International Conference on Cloud Computing*, Santa Clara, 2013.
[38] C. Nagaraj and N. M. Sathish kumar, "Cloud Computing With a Model Futuristic Maturity," *International Journal of Advance Research In Science And Engineering,* 2015.
[39] C. Guangming , L. Yao , G. Zhiwei and L. Xiaoyin , "Cloud data governance maturity model," in *8th IEEE International Conference on Software Engineering and Service Science*, Beijing, 2017.
[40] G. Conway, D. E. Doherty, D. M. Carcary and C. Crowley, "Enterprise Cloud Adoption - Cloud Maturity Assessment Model," in *ECISM 2017 Conference Proceedings publication*, Maynooth, 2017.
[41] S. Alqassemi, Y. K. Ever and A. V. Rajan, "Maturity Level of Cloud Computing at HCT," in *Fourth HCT Information Technology Trends*, Dubai, 2017.
[42] S. G. Grivas, M. Peter, C. Giovanoli and K. Hubli, "FHNW Maturity Models for Cloud and Enterprise IT," *Business Information Systems and Technology 4.0,* pp. 133-146, 2018.

An Information System Design Approach Suitable to the Social Contexts of Japan

TETSUYA UCHIKI
Graduate School of Humanities and Social Sciences, Saitama University
*eMail:*uchiki@kyy.saitama-u.ac.jp

ABSTRACT

The way of recognition and design of IS change greatly with the cultural environment. Especially in Japan, although the requirements of users are researched, only the shallow problems dependent on the current system are solved. And the social or the organizational states of the machinery system and the functionality of the users is not argued as the core of the system design. Therefore, many of the IS have not been designed including the real activities of the users who are the main players in the system. It is thought that many social problems involving IS have been caused by deviance of users induced under this situation. This paper focuses to the incidents of security violation about the personal information as an example of the IS problem of Japan resulting from a system design. And the IS problems, such as the functionality of ICT, consciousness of systematic governing structure and a user, and morals, are analyzed from the viewpoint of suitability for the cultural environment. Finally, the approach method of a design suitable to cultural environment of Japan is discussed.

❖ IS Design Approach Suitable to Social Context of Japan

Introduction

Generally, an information system (IS) is recognized through the use of a physical ICT system because the presence of IS is strongly impressed as practical experience. Therefore, in design and development of IS in the early era, the realization of ICT systems which can perform the target processing certainly was the first priority [1]. Under expansion of the application domain of the ICT system and diversification of its usages accompanying progress of the information technologies, the definite usability of ICT and the suitability to a user's environment became important as a design of IS. And then, the various methodologies for designing IS suitable to the environment of the users have been proposed and used [2]. At present when standardization and rationalization of the technical system progressed, designing the function of users in IS as main players is the key factor of success of a system design.

The way of recognition and design of such IS, however, change greatly with the cultural environment. In Europe consisting mainly of Northern Europe or UK, the design of a logical system based on people's acts and the meanings of information to treat is thought as important. In the U.S., the design aims to develop the infrastructure of information executable the corporate activity adapted to a business model rationally and smoothly. Also, in Japan, original design methodology has been advocated and employed by the vendors of ICT or SI under the same social context. However, there are few the scientific arguments and researches about IS designs, such as the usage and the problems, also we cannot find even a textbook about IS. Although many design methodologies have been introduced in the society and the research groups, most of the researchers were only interested in the concrete functionality. The argument about the essential use of the methodologies was hardly made. There are few business schools which lead the design and modeling of IS, and also the education about IS has not been performed in general. On the other hand, the graduate schools of the ICT specialist for the purpose of technical problem-solving approach for IS design have been established.

Of course, many IS problems have occurred in Japan, as well as other cultural environment. The adaptability of the work field and a dexterity of user are reason why the IS design has not been recognized as a social problem in spite of this situation. Actually, a "KAIZEN" in which users invent dexterous ways to perform their works, even EUD which develops the user-oriented systems by themselves are observable at the work fields of a ICT system. IS design is not visible and the scale of quality is not clear. Most of the users have been able to use the ICT systems without their complaints. A developer wants to evade the process for the design which requires time and time and effort. It is thought that the social situation of IS designing only about functionality was formed by these causes.

In Japan, however, although the requirements of users are researched, only the shallow problems dependent on the current system are solved. And the social or the organizational states of the machinery system and the functionality of the users is not argued as the core of the system design. Therefore, many of the ISs have not been designed including the real activities of the users who are the main players in the system. It is thought that many social problems involving an IS have been caused by deviance of users induced under this situation. At the era when most of ISs were used in the closed organization limitedly, the problems of the design have been little revealed. Since the systems, which connected always to the Internet, have come to be used by various users in the global environment, the problem involving the design has been actualized. Especially, although most of them were not serious, the number of the incidents or the accidents involving the personal information outflow of their customers or users which the organizations or the companies have collected, are remarkable. Each phenomenon has been individually recognized as a problem of technology, management, or a user's consciousness, and has been solved separately. However, most of the phenomena should be caught as the problem of the system design which includes each factor fundamentally

This paper focuses to the incidents of security violation about the personal information as an example of the IS problem of Japan resulting from a system design. And the IS problems, such as the functionality of ICT, consciousness of systematic governing structure and a user, and morals, are analyzed from the viewpoint of suitability for the cultural environment. Finally, the approach method of a design suitable to cultural environment of Japan is discussed.

Security Incidents as IS Design Problems

Although there are many types of the security violation incidents about the information held inside the ICT system, the incident of flowing out of personal information attracts attention particularly. Japan Network Security Association (JNSA) as NPO reported that 1000 or more incidents about the personal information have occurred in every year from 2005 which began to take statistics to 2015, as shown in Table 1 [3]. The main point is which most of those incidents were using a paper medium, as shown in the graph of Figure 1. Since the practical use of information is not easy and the volume size of the violated information for every incident have been comparatively slight in a paper medium, it was thought that most of the incidents have not been a critical social problem. Reduction of the incidents from 2015 can be also thought the reflection of the use of a paper medium decreasing. In current society where the ICT system progressed, it is necessary to pay attention to using many paper media.

The causes of a flowing out of personal information can be divided roughly into two types. One is unlawful access of the outsiders, and other is the incidents of insiders. The insiders' incidents are misoperation and carrying out data. As shown in Figure 3, in Japan, it turns out that the latter incidents occupy the most. Most of the cases are the incidents which flow out by the printed documents, USB memories, or E-mail which recorded information for business use. They are not intentional acts but the mis-management or the mis-operation on a performing their works. Especially it is important that the printed documents, PCs, or memory devices using in their business processes have been lost, stolen. It is clear that the system designs, which can print out or download to local PCs such critical information to do the operations of ordinary business, are problems. And they were caused by the users' mistakes in management or operations of ordinary business.

On the other hand, it is clear that the system designs, which must carry out such critical information to execute the ordinary business operation, are problems. The systems should be designed including the business process re-engineering so that the acts which induce mistakes and troubles may be eliminated originally. The users understanding the system operations deeply, and having the height of morals and motivation are the reason why the system which included such a design problem could be used in the business practically without a grave fault.

Table 1 Reported Security Incidents in Japan

Year	Reported Incidents	Victims
2005	1,032	8,814,735
2006	993	22,236,576
2007	864	30,531,004
2008	1,373	7,232,763
2009	1,539	5,721,498
2010	1,679	5,579,316
2011	1,551	6,284,363
2012	2,357	9,720,065
2013	1,389	9,254,513
2014	1,591	49,999,892
2015	788	4,956,953
2016	468	13,965,227

❖ IS Design Approach Suitable to Social Context of Japan

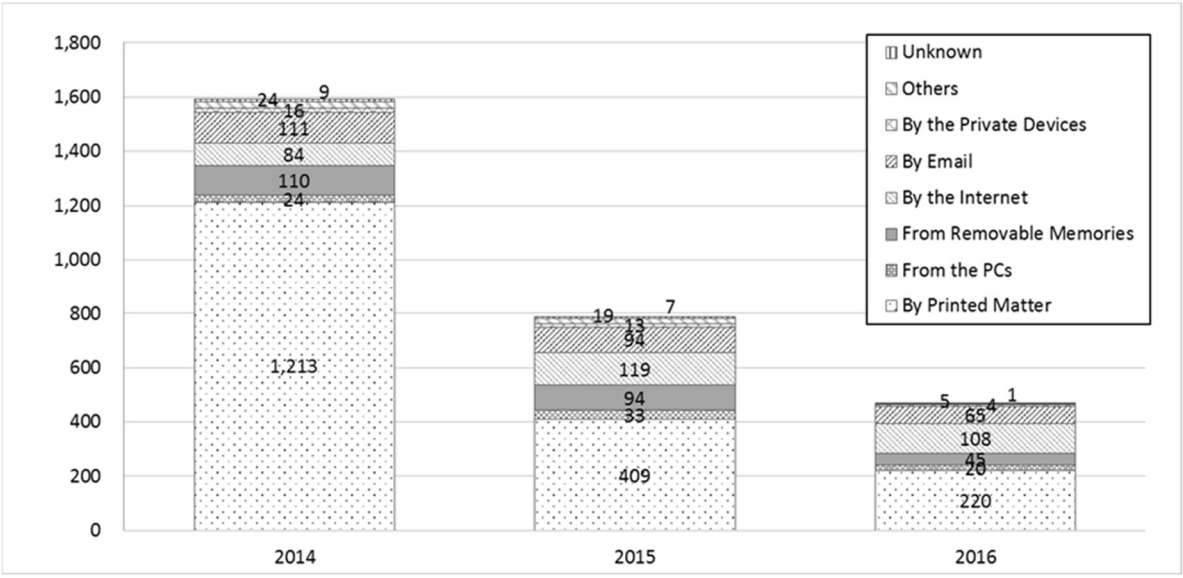

Figure 1 The Media Used in the Reported Incidents

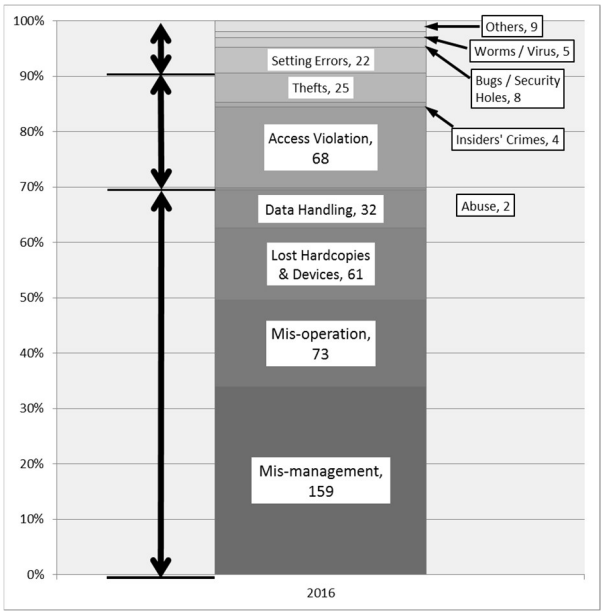

Figure 2. The cause and its frequency of the incidents reported in 2016

Analysis of Social Context of IS Design

The recognition and the consciousness of IS design are changed greatly with the roles and positions in the society of the users. Analysis of the social context of IS design in Japan The recognition and the consciousness of IS design are changed greatly with the roles and positions in the society of the users. Hirschheim, R. and et al. classified the design stance which differs in the recognition, into four paradigms based on the classification axis of the sociological paradigms proposed by G. Burrell & G. Morgan [4][5]. Hirschheim's framework can be regarded as four different viewpoints of IS [6]. It clarifies the difference of the recognition of IS by the situation of user in the society, as shown in Figure 3.

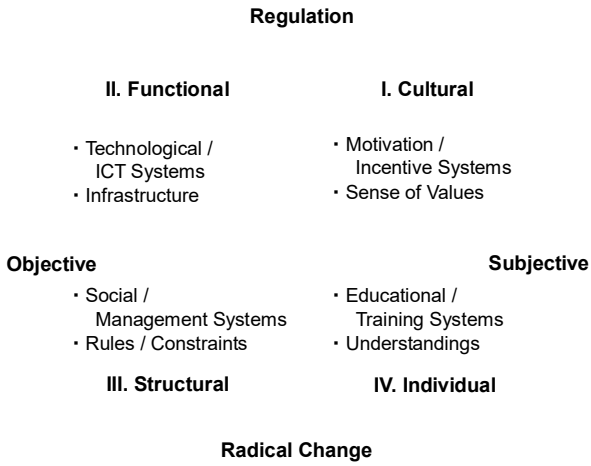

Figure 3: The difference of IS views by roles or positions in organization

A technological solution policy is only one viewpoint of the four paradigms of matters related to activate IS. It means that the solution needs the measures for other three mat-ters simultaneously. It is required for development of IS to plan the policy for consid-

ering and forming the state of the organizational constraints, the social culture, and the users' consciousness to introduce the system. For IS to keep on working actively, the system itself always must offer the environment where the users who are its main constituent can work smoothly and efficiently. However, the system as a technological solution can not only hardly offer such environment, but also not bring out the capabilities of the system if the users' consciousness are not so high. Moreover, IS is not activated probably under the social context as which constraints and the regulations of system usage are rigorous or where the sense of using IS is not evaluated. Thus, in order to activate the functionality of IS, the system has to be designed as technological solution in consideration of the balance among social constraints, a norm, and user consciousness, as shown in Figure 3. And in the operation and the maintenance of IS, it is important to keep the balance of them.

The implementation of new ICT always includes not only the effect expected but also a certain side effect. In many cases, the effects emerge slowly including even a time lag indirectly rather than directly. The processes of IS development can be divided into a process related with a functional ICT system, and another process related with the social system and culture, as shown in the two ellipses of Figure 3. And the relation of two processes can be shown in Figure 4. The developed ICT system will change not only the users' functionality of information processing, but also behaviors of the users. Such changes of users affect the ordinary behaviors in every life and also the ICT system will be required as an infrastructure in the society. This dynamic change of society is considered to be a re-construction process of the environment involving an ICT system, like known as the business process re-engineering (BPR) in corporations. These processes evolve the structure of a society to the IS created by the ICT system contrary to the development process of an ICT system. And also the processes might be required from the actual users of the ICT system in their life, as bottom-up process.

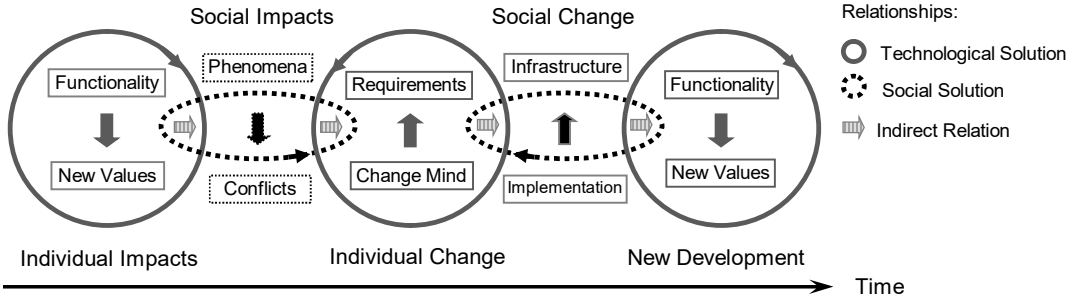

Figure 4: Dynamics of the Social Impacts of ICT System Developments

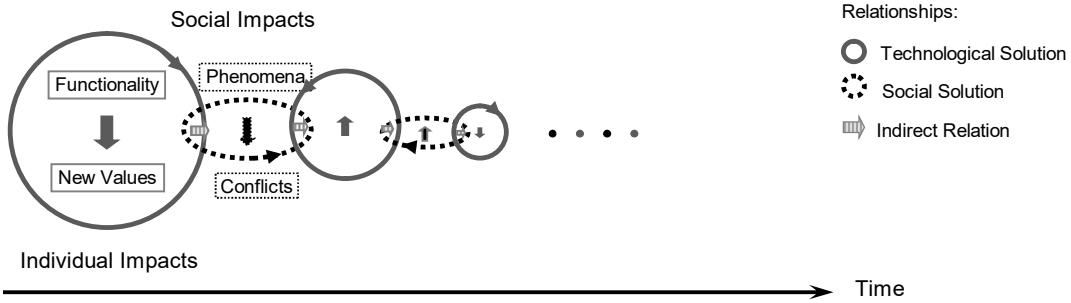

Figure 5. Dynamics of IS Design Cycle in Japan

❖ IS Design Approach Suitable to Social Context of Japan

In Japan, however, it is rare to clarify the problems and demands that a user points out the mis-match design of the introduced ICT system, and / or a fatal trouble arises to in business fields. It is thought that the reason originates in having the dexterity to which many users compensate for the problems by own operation, and having the "KAIZEN (improvement)" consciousness to which the usage is fitted to their own needs [7]. Therefore, in Japan, a cycle as shown in Figure 4 which fixes new infrastructure and is continued expansively is not activated. On the contrary, as shown in Figure 5, it seems that the influence of the introduced ICT system is absorbed, and resolved the problems.

The essential design of IS is not so important in such cultural environment of Japan, and also the effective design methods have not been developed. The essential design of IS is not so important in such cultural environment of Japan, and also the critical design methods have not been developed. Moreover, in the social context of Japan, the social recognition for professionals like a system designer is not good. Therefore, their consciousness to a duty as professionalism is not only to design the required systems, but also they have to present their skill and knowledge. The system designers as a professional under Japanese social context could not design the critical systems which needs various cooperators in the client organizations, so they had to design the effective systems depending on the legacy environment realizable only by engineers.

In the cultural environment as shown in Figure 4, IS designer will take part in an administrative activity in the society to give requirements from the user's situation. And the role will expand to deciding upon a policy of the system contributed to construction of new ICT systems as the infrastructure in the society as CIO. On the other hand, under the environment of Figure 5 like Japan, since a designer's activity will be included in the system performance, IS designer, as a functional element indispensable to an administrative action, will be only ordered to realize the functional system from top management. In Japan, the social context involving IS has formed the social situation which cannot rear truly trained CIO and does not demand them.

Consideration on IS Design in Japan

Thus, the design method of IS under the quite different cultural environment from Western countries, must be researched uniquely. We have considered that both the nearness of the distance of users and professionals and the high level adaptability of users are the key factor. If the designers employ the design methods from the U.S. under the social context of Japan, they have to make the users conscious of distance and specialties as a technological skill, and their slight authorities cannot enlighten the management enough. Therefore, since the system designers become like a manager's spokesman, they are isolated from both of the users and the management. In Japan, a system designer should aim at becoming a spokesman of user taking advantage of the nearness of distance with users and a design methodology for such systems is needed.

IS design is a development policy of the ICT system, which can harmonize with the social context, and must make a decision politically. Thus the great difference on the IS design dependant on the social context in each cultural environment is naturally. As mentioned above, however, the social context of IS design in cultural environment of Japan is characteristic. It is thought that such differences arise from an adaptation ability and a behavioral trait of users. As shown in Figure 4, the goal of IS design is a social infrastructure design in Western cultural environment. The users act freely with their maximum capability. It is the same as playing original performance on the standard stage as shown in Figure 6a. On the other hand, in Japan, each user requires original IS design according with own activity. Although each one seems to play on the original own stage as shown in Figure 6b, the performance is standard. It is originating in a Japanese user's dexterity. Rather than performing freely on the regular stage, a dexterous user is considered to desire to perform standard performance on the original own stage, in order to impress with own dexterity. If so, Japanese IS design should aim to offer the activity infrastructure which can choose freely the performing methods and tools, although IS design in Western cultural environment is specified as a free activity infrastructure.

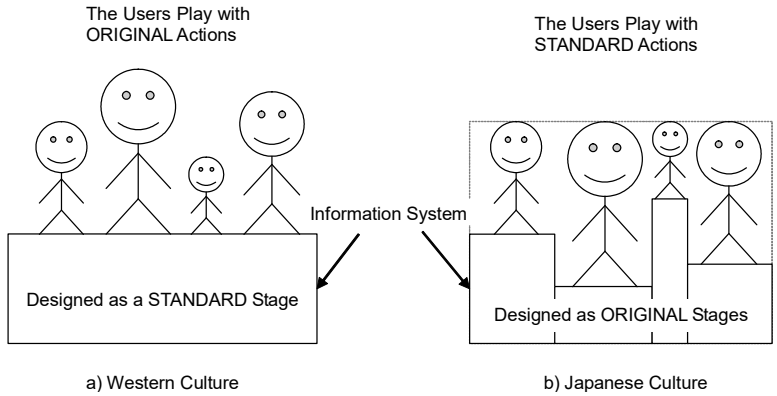

Figure 6. The different situations of the IS in the Western and Japanese culture

Concluding Remarks

Since the social context involving IS design is deeply dependent on cultural environment, it cannot be expected that such a situation accomplishes a drastic change according to the design methodology from other cultural environment. Although a strong conviction and the strong vitality are indispensable to a design activity, the danger of falling self-righteous cannot be denied. It can be thought that the cultural environment of Japan have performed refusing it. By the academic researches of IS design suitable for the user of Japan where recognition and consciousness of IS differ from a Western user, it is thought that the deep knowledge which contributes to an essential understanding of IS and the way of approach are acquired.

References

[1] Uchiki, Tetsuya (2014): The Cultural Feature of Japan in the Historical Transition of Information Systems Design, *SEFBIS Journal,* No.9, GIKOF, Hungary, pp.2-7.

[2] Avison, David and Fitzgerald, Guy (2006): *Information Systems Development (4th Ed.),* McGraw-Hill Education (UK) Ltd.

[3] NPO Japan Network Security Association (2017): *2016 Security Incident Survey Ver1.2,* http://www.jnsa.org/result/incident/data/2016incident_survey_ver1.2.pdf.

[4] Hirschheim, Rudy, Heinz K. Klein, Kalle Lyytinen (1995) Information Systems Development and Data Modeling, Cambridge University Press.

[5] Burrell, Gibson and Morgan, Gareth (1979): *Sociological Paradigms and Organizational Analysis,* Heinemann.

[6] Uchiki, Tetsuya (2009): The Characteristics of Researches on Enterprise Information Systems in the Social Context of Japan, *Proceedings of CONFENIS2009,* Györ, Hungary, 2009/10.

[7] Nonaka, Ikujiro and Takeuchi, Hirotaka (1995): *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation,* Oxford University Press.

The Impact of Industry 4.0 to the ERP Approach

[1]MUHAMET GËRVALLA . [2]KATALIN TERNAI
[1] University for Business and Technology, Kosovo, [2] Corvinus University of Budapest
E Mail: muhamet.gervalla@ubt-uni.net, ternai@informatika.uni-corvinus.hu

ABSTRACT

Industry 4.0 is seen as the beginning phase where computers and automation become connected and as an opportunity to increase the efficiency and effectiveness in the manufacturing industry. With the application of real-time data and information by integrating physical machinery and devices with networked sensors and software to predict, control and reduce costs in a long-term view. This study presents the basic fundamental concepts of Industry 4.0, where each technology of Industry 4.0 is described based on the current literature. Also, this study tries to understand the impact of Industry 4.0 to the ERP approach and how each technology of the Industry 4.0 can be related to ERP system and if current ERP systems are ready to support Industry 4.0. Based on the findings, there are many challenges related to the integration of Industry 4.0 and current ERP systems especially when it comes to the machine to machine, machine to ERP communication and the security of the data.

Introduction

The business environment in most industries is becoming more complex and more internationalized, which means that managing and using information effectively is very important for the success of modern companies. Managing all operations in a company requires information systems that can integrate data starting from managerial to production (entire Product Lifecycle Management) and available for use at any time and accurately. Many companies are using solutions that IT provides, such as ERP (Enterprise Resources Planning) systems, to manage their company processes, and to integrate all the various operations in order to increase the flow of information within the company as well as collaboration with partners, suppliers, and customers.

ERP systems help companies in many areas starting from increasing better information sharing between departments, improving workflow, better supply chain management, integrating of data, processes, and technology in real-time across internal and external value chains, standardization of various business practices, improve orders management and accurate accounting information of inventory management [1]–[3].

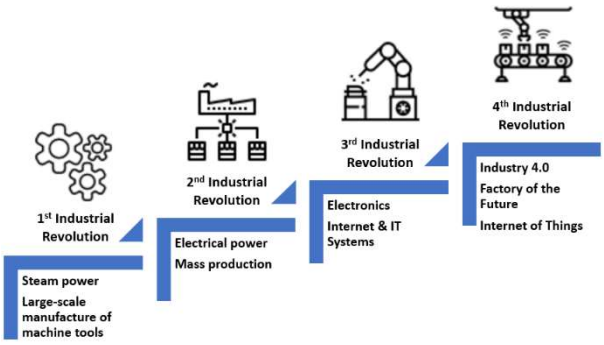

Figure 1. Industrial revolution adopted from Haddara et al. [4]

On the other hand, industrial production started with the 1st generation where machines were mechanized production by using water and steam; 2nd generation started with mass production by using electricity; 3rd industrial revolution came about the use of computers and electronics to automate production. And now, we have the 4th generation or Industry 4.0 which is in the beginning phase where computers and automation become connected by using IoT devices, starting from suppliers, distributors, manufacturing facilities and when robots and machines began to replace human workers [4], [5].

Methodology

The aim of this paper is to draw a picture of the current state of Industry 4.0 and its relation to ERP systems. The study was done by obtaining and analyzing existing research through secondary data (published papers, academia, industry, and other relevant credible sources) to review through and to present the revolution of Industry. The aim of this study is: 1. to identify the technologies of Industry 4.0 through a comprehensive and systematic literature review; 2. to understand the relationship of the Industry 4.0 with ERP systems; and 3. to identify challenges to integrate Industry 4.0 technology and ERP systems. Through the reviewed literature, it has been achieved to meet the objectives of this research and to reach a conclusion in reference to the impact of Industry 4.0 on the ERP systems approach. Although much work remains in future studies to understand the readiness of ERP vendors to support Industry 4.0 in their ERP systems.

The concept of Industry 4.0

Industry 4.0 is the fourth industrial revolution which presents the transformation and application of new and innovative technologies in the manufacturing sector. This concept uses the internet as a mechanism to connect factories in many fields. Below are presented some different definitions according to different researchers.

According to the German Federal Ministry of Education and Research, the term Industry 4.0 has existed since 2011 when strategic manufacturing roadmap was developed to promote the digitalization of manufacturing. Industry 4.0 is also called 'smart industry', 'intelligent industry', 'smart factory' or 'smart manufacturing' [6].

"The implementation starts with small steps here and there, there won't be a big bang that is going to introduce Industry 4.0. On the contrary, it will come step by step. But if we look back in ten years, we will see that the world has changed significantly." (Hartmut Rauen, Deputy Executive Director Mechanical Engineering Industry Association - VDMA, 2012) [23], [24]. "Industrie 4.0 is the integration of complex physical machinery and devices with networked sensors and software, used to predict, control and plan for better business and societal outcomes" [7].

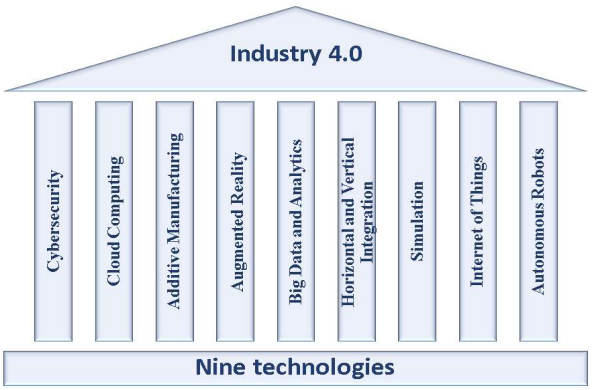

Figure 2. Industry 4.0 technologies

Industry 4.0 as a new concept, to fulfill the modern manufacturing requirements applies nine technologies which are presented in Fig. 2. These technologies allow more flexibility in production and real-time monitoring, controlling and reaction based on the real-time situation and requirements.

Industry 4.0 aims to reduce the complexity of operation in manufacturing to increase the efficiency and effectiveness by application of real-time data and information which are interconnected by IoT sensors in order to reduce costs in long-term for companies [8]. Below are presented some technologies of the Industry 4.0.

Cybersecurity

Shafiq et al. define CPS - Cyber-Physical System as the established global network which is implemented in global networks for business which is combined with physical and digital worlds that include: warehousing systems, machinery, and production facilities [7]. Monostori and Váncza definition of CPS is: "Cyber-physical systems are assembled of collaborating computational entities which are in intensive connection with the surrounding physical world and its ongoing processes, providing and using, at the same time, networked data accessing and data-processing services available typically on the Internet" [9].

According to Gandhi from SAP, CPS is able to adapt to dynamic requirements and therefore is self-optimizing. This helps in automatization and decentralization of processes in collaboration network, with machines, products, objects, warehousing system and production facilities [10].

Cybersecurity is a crucial point in Industry 4.0, because of the increased number of devices which are interconnected. Industry 4.0 requires unified standards and communication protocols. Many devices are used in Industry 4.0, starting from machine controllers, sensors, manufacturing lines, and other industrial systems, so the cybersecurity threats will increase dramatically [11]. It is very crucial to ensure that all communications equipment and protocols are secured to protect critical systems from cybersecurity threats.

Simulation

Simulation is the tool for predicting and evaluating the performance of systems that are analytically intractable. By integration of sensing, computing, and control, Jie Xu et al. defines that simulation optimization helps companies in the decision-making process which provides the "smart brain" required to drastically improve the efficiency of industrial systems [12].

With the use of future simulations, companies are enabled to simulate the real-world situation in a virtual model, which can help companies to enable testing and optimizing of products, places, etc. in the virtual world before the physical setup.

Horizontal and Vertical Integration

Vertical integration is focused in the connection of different levels in company with the help of IT systems, especially in production management, manufacturing, and low-level PLC (Programmable Logic Controller) systems like machine controllers, sensors, etc. that exist within the company in order to increase the flexibility and performance in planning and management [13].

Horizontal integration implies the connection between all the components of the value chain, starting from internal company logistics, production, sales and services, to external partners, suppliers, customers, energy suppliers, etc. to create a value chain as autonomously acting participants [13].

Cloud Computing

Cloud computing is defined as an Internet-based service or IT infrastructure starting from applications delivered as a service or hardware and software in the data centers provided by a service provider that is always available [14]. Cloud computing can be divided into three categories [15].

1. Software as a service (SaaS): is a model of software where a provider licenses an application which is delivered over the internet. SaaS providers host applications on their web server and simplify the utilization of a large number of software applications remotely, elastically and seamlessly [16];
2. Platform as a service (PaaS): a software development framework and components all delivered on the network. PaaS model is a platform which includes an operating system, programming language execution environment, database, and web server. A PaaS client is able to develop and run its applications at the software layer [16];
3. Infrastructure as a service (IaaS): an integrated environment of computing resources, storage, and network fabric delivered over the network. Offered as an on-demand, pay for usage model [16].

Additive Manufacturing

Additive Manufacturing is a technology which allows companies to produce a prototype, individual components, and 3-D printing. "With Industry 4.0, additive-manufacturing methods will be widely used to produce small batches of customized products that offer construction advantages, such as complex, lightweight designs. High-performance, decentralized additive manufacturing systems will reduce transport distances and stock on hand" [11].

Augmented Reality

Augmented Reality (AR) can be defined as the ability to combine the physical real-world environment information by adding virtual information which is generated by computers [17], [18]. Yang defines AR as the technology of the future, who can develop "next generation, reality-based interface" [19]. Augmented Reality has many advantages compared to Virtual Reality. The main advantage of AR is the ability of integration virtual environment and the real-world interaction [19]. By using the Augmented Reality, the user can increase capability of finishing tasks by using virtual information from different

sources directly to his work environment like the live-video streaming, or just getting the instruction how to operate with different kinds of equipment even if the technician is not an expert in that particular part of the equipment which is presented in SAP and Vuzix cooperation [20]. Augmented Reality architecture has four main steps in his process: scene capture; scene identification for choosing the accurate information for boosting it; scene processing and visualization of the augmented scene [21]

Augmented Reality based systems nowadays are used in different aspects in real time situation, starting from warehouses, maintenance instructions, etc. which helps the users of this technology to improve decision making and work procedures [11].

Internet of Things

Bacsárdi & Gludovátz, declare that there are many reasons to apply the Internet of Things in the Industrial field: "now: the companies can reduce the cost of operation, and prevent the failure or stoppage of the production line −in the future, the companies gain extra profit via service-oriented production system and the managers' needs will be satisfied for easier decision making" [27].

The application of Internet of Thing devices can contribute to the data reading and transferring to the main databases, in the same time these equipment's allow automatization of the data entry, which helps in reduction of data entry errors and data processing time.

Big Data and Analytics

Big data can help companies to analyze the past, present and predict the future, by using an analytical application to generate value from the available data, based on the four V's of Big Data: Volume, Variety, Velocity, and Veracity [22]. Bacsárdi & Gludovátz, mentioned in their paper that it's very important to use the big data in the right way [27]. They described the usage of Big Data in two steps. Initially, *data management*, which includes data collecting, storing and preparing for analysis, and *analytics*, the tools which are used for Big Data Analytics to get the smartest solution on a process. According to them, big data are useless if they are not analyzed. Below are presented some facts from an IBM article: Extracting business value from the 4 V's of big data, as they are presented below [22]:

- *Volume (Scale of data):* 90% of today's data has been created in the last two years. Every day we create 2.5 quintillion bytes of data;
- *Velocity (Speed of data):* Every 60 sec there are 72 hrs. of footage uploaded to YouTube, 216000 Instagram posts, or 204000 emails sent. 50T GB/sec is the estimated rate of global Internet traffic by 2018;
- *Variety (Diversity of data):* 80% of data is video, images, and documents and 90% of them are unstructured (tweets, photos, etc.);
- *Veracity (Certainty of data):* 1 in 3 business leaders don't trust the information they use to make decisions [22].

Industry 4.0 and ERP integration

Hochmuth et al. on a Deloitte report, emphasize that the role of ERP systems will change, from a central database system which collects data, to support mobile role-based user interactions [25]. They declare that if ERP system vendors want to support Industry 4.0, they need to adapt their ERP systems to the technical and process-related requirements, in context of data storage, data exchange, and data usage.

Haddara et al. in their research raised the question if ERP systems are ready for the FoF (Factory of Future) [4]. According to them, within this new concept of FoF, it is required a real-time two-way communication between machines, processes, and products. Here comes the question if ERP systems can support such communication, which will bring ERP to the next level.

Porter and Heppelmann in their article How Smart, Connected Products are Transforming Competition, presented an analysis of the impact of IT on the competition waves [26]. Initially, they give an overview of the first wave of IT, which started with automation in the manufacturing industry, with MRP (Manufacturing Resource Planning) and MRP II technologies. Then, they continue with the second wave, where the internet was viewed as an opportunity for doing business, by enabling coordination

and integration between different levels across enterprises in focus to the relation of suppliers and customers in a global market where the Internet was seen as a mechanism for information sharing. This wave, in other words, presents the ERP systems application. According to Porter and Heppelmann, in the third wave of competition, operational effectiveness is the key to competition, and smart connected products can contribute to it by changing the approach to product design, services, marketing, human resources, and security. Authors in this article identified four dimensions, how the usage of smart connected products can contribute.

- *Monitoring, real-time information related to the operation of products reported by them self.*
- *Control*, the ability to manipulate, maintaining the machines from a distance.
- *Optimization*, managing the resources by using data generated from real-time monitoring and optimizing the product operation and production capacity.
- *Autonomy*, adoption to dynamic requirements for self-optimizing, coordination, and diagnosis. According to the authors above, can be concluded that the future of ERP systems is closely related to Industry 4.0 technologies.

In theory, Industry 4.0 contains nine technologies and each of them is related to the ERP systems somehow, some of them are analyzed below. Starting from the Internet of Things (IoT) which is the first technology of Industry 4.0 which should be supported by ERP systems. With the support of IoT, all other Industry 4.0 technologies could be interconnected with ERP systems. IoT should be an integration point of Industry 4.0 and ERP systems. The integration of this technology in Industry 4.0 is closely related to communication protocols between the machines and different frameworks which should be harmonized in the future.

Nowadays, in most of the cases, data from manufacturing warehouses, production planning, quality control, processes, and other sources are entered manually in ERP systems. The use of IoT would help to automatically enter data into ERP systems which are collected from different production sources and this would have the effect of eliminating data entry errors. Entering data through the use of IoT equipment would also help in changing the approach to ERP systems. There would be another link between ERP and Big Data Analytics and the other technologies of Industry 4.0. With the use of Big Data Analytics, ERP systems can enable collection and evaluation of data from different sources in real time and can help in decision making, quality control, optimizing costs and other aspects. Also, Simulation the other technology of Industry 4.0, can relate to ERP through the use of Big Data Analytics for prediction and evaluation of the performance of systems that are analytically intractable to simulate in order to test and optimize the resources.

With the use of ERP system which can be supported by Big Data Analysis in combination with technologies like smart glasses can be applied Augmented Reality which can completely change the way how maintenance services and performing works in the warehouse can be achieved. After the scene has been captured by smart glasses it can be identified by using data that are stored in ERP system and choosing the accurate information for boosting it for scene processing and visualization of the augmented scene. Also, employers can finish their tasks by using these technologies which support of 3D Visual Enterprise models of workplace, getting instruction from ERP systems, finding products in the warehouse, checking if the products are available in the stock and all the information directly to the smart glasses which help the employers to make decision based on the information they are getting form ERP systems. By using Horizontal and Vertical integration approach supported by ERP systems autonomous production methods can be powered by using Robots that can complete tasks intelligently, with the focus on safety, flexibility, versatility, and collaborative.

Different ERP system vendors should think about integrating and supporting smart connected products into their systems by applying Industry 4.0 technologies. This integration would be a benefit in generating value for customers, creating a new competitive environment, increasing company productivity, and global economic growth.

Conclusion

This study is a descriptive analysis in the ICT discipline where argues that one of the major concerns about Industry 4.0 and ERP systems integration is whether the current ERP systems are ready for Industry 4.0. The benefits of the integration of Industry 4.0 and ERP systems, will be very high, especially in support of intelligent machines and processes, data analytics and modeling which can help companies in the real-time decision-making process.

Based on the concepts analyzed in this study, ERP systems can be easily integrated with Industry 4.0, but there are also some challenges when it comes to M2M (machine to machine) and machine to ERP communications because there are no unified standards and protocols [4], [11].

Another important issue in Industry 4.0 is security. This is due to the diversity of devices and technologies that are interconnected and a large amount of data passing through various communication channels.

ERP systems should support Cyber-Physical Systems which leads to Feature end-to-end vertical integration, from inbound logistics, planning, marketing to outbound logistics, and services. Also, ERP system providers should integrate the MES - Manufacturing Execution Systems into their application, to create a new approach on ERP systems that could be implemented in manufacturing industries that can have a large impact and application in Industry 4.0.

References

[1] S. Shang and P. B. Seddon, "Assessing and managing the benefits of enterprise systems: the business manager's perspective.," Inf. Syst. J., vol. 12, no. 4, pp. 271–299, 2002.

[2] D. L. G. Thomas F. Gattiker, "What Happens after ERP Implementation: Understanding the Impact of Interdependence and Differentiation on Plant-Level Outcomes," MIS Q., vol. 29, no. 3, pp. 559–585, 2005.

[3] A. V. Mabert, A. Soni, and M. A. Venkataramanan, "The impact of Organization Size on Enterprise Resource planning (ERP) Implementation in the US Manufacturing Sectors," Int. J. Manag. Sci., vol. 31, no. 31, pp. 235–246, 2003.

[4] M. Haddara and A. Elragal, "The Readiness of ERP Systems for the Factory of the Future," Procedia Comput. Sci., vol. 64, pp. 721–728, 2015.

[5] Forbes, "What Everyone Must Know About Industry 4.0," 2016. [Online]. Available: https://www.forbes.com/sites/bernard-marr/2016/06/20/what-everyone-must-know-about-industry-4-0/#70e537ad795f. [Accessed: 10-June-2019].

[6] Business Insider, "How the Internet of Things is revolutionizing manufacturing," 2016. [Online]. Available: http://www.businessinsider.com/internet-of-things-in-manufacturing-2016-10. [Accessed: 10-June-2019].

[7] S. I. Shafiq, C. Sanin, E. Szczerbicki, and C. Toro, "Virtual engineering object/virtual engineering process: A specialized form of cyber physical system for industrie 4.0," Procedia Comput. Sci., vol. 60, no. 1, pp. 1146–1155, 2015.

[8] C. Santos, A. Mehrsai, A. C. Barros, M. Araújo, and E. Ares, "Towards Industry 4.0: an overview of European strategic roadmaps," Procedia Manuf., vol. 13, pp. 972–979, 2017.

[9] J. Váncza and L. Monostori, "Cyber-physical Manufacturing in the Light of Professor Kanji Ueda's Legacy," Procedia CIRP, vol. 63, pp. 631–638, 2017.

[10] N. Gandhi, "Industry 4.0 – fourth industrial revolution," 2015. [Online]. Available: https://blogs.sap.com/2015/06/30/industry-40-fourth-industrial-revolution/. [Accessed: 10-June-2019].

[11] M. Rüßmann et al., "Industry 4.0. The Future of Productivity and Growth in Manufacturing," Boston Consulting, 2015. [Online]. Available: https://www.bcg.com/publications/2015/engineered_products_project_business_industry_4_future_productivity_growth_manufacturing_industries.aspx#chapter2_section7. [Accessed: 10-June-2019].

[12] J. Xu, E. Huang, L. Hsieh, L. H. Lee, Q. S. Jia, and C. H. Chen, "Simulation optimization in the era of Industrial 4.0 and the Industrial Internet,"

J. Simul., vol. 10, no. 4, pp. 310–320, 2016.
[13] ICA, "Industrie 4.0 Controlling in the Age of Intelligent Networks," Int. Controll. Assoc., pp. 1–56, 2015.
[14] M. Armbrust et al., "A view of cloud computing," Communications of the ACM, 2010. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1721654.1721672.
[15] W. Wu, L. W. Lan, and Y. Lee, "International Journal of Information Management Exploring decisive factors affecting an organization ' s SaaS adoption : A case study," Int. J. Inf. Manage., vol. 31, no. 6, pp. 556–563, 2011.
[16] X. V. Wang and X. W. Xu, "An interoperable solution for Cloud manufacturing," Robot. Comput. Integr. Manuf., vol. 29, no. 4, pp. 232–247, 2013.
[17] H. K. Wu, S. W. Y. Lee, H. Y. Chang, and J. C. Liang, "Current status, opportunities and challenges of augmented reality in education," Comput. Educ., vol. 62, pp. 41–49, 2013.
[18] J. Carmigniani, B. Furht, and M. Anisetti, "Augmented reality technologies, systems and applications," … Tools Appl., 2011.
[19] R. Yang, "The study and improvement of Augmented reality based on feature matching," 2011 IEEE 2nd Int. Conf. Softw. Eng. Serv. Sci., pp. 586–589, 2011.
[20] SAP&Vuzix, "SAP and Vuzix bring you the future of Field Service," 2014. [Online]. Available: https://www.youtube.com/watch?v=UlpGDrSmg38. [Accessed: 10-June-2019].
[21] A. O. Alkhamisi and M. M. Monowar, "Rise of Augmented Reality: Current and Future Application Areas," Int. J. Internet Distrib. Syst., vol. 01, no. 04, pp. 25–34, 2013.
[22] IBM, "Extracting business value from the 4 V ' s of big data," 2016. [Online]. Available: http://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data. [Accessed: 10-June-2019].
[23] Martinez, Felipe, and Richard Brunet-Thornton. Analyzing the Impacts of Industry 4.0 in Modern Business Environments. Business Science Reference, 2018.
[24] E. Hofmann and M. Rüsch, "Industry 4.0 and the current status as well as future prospects on logistics," Comput. Ind., vol. 89, pp. 23–34, 2017.
[25] Hochmuth, A. C., Bartodziej, C., & Schwagler, C. (2017). Is your ERP system ready for the digital era?, 50. [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/de/Documents/technology/Deloitte_ERP_Industrie-4-0_Whitepaper.pdf [Accessed: 02-June-2019].
[26] Porter, M., & Heppelmann, J. (2014). How Smart, Connected Products are Transforming Competition. SPOTLIGHT ON MANAGING THE INTERNET OF THINGS, Harvard Business Review.
[27] L. Bacsárdi and A. Gludovátz, "Industry 4.0 projects' background and our experiences at the wood industrial manufactories," SEFBIS Journal., vol. XI, no. March, pp. 34–41, 2017.

A New Approach and Framework for Risk Assessment Maturity

BLERTON ABAZI
University for Business and Technology (Kosovo) Faculty for Information Systems
eMal: blerton.abazi@ubt-uni.net

ABSTRACT

Recently, we have had a huge increase in the use of cloud services, so the gap between internal and external networks almost disappear. Trust in the cloud will increase in the possibility of more attacks in the cloud. Adding to this the entrance of many new digital devices in the market, has made it difficult to use the traditional method of data protection. For this reason, companies in business and public sector must deal with risks management, develop and maintain the related processes for their systems. Companies that are responsible for information

security, continuously develop systems and update the existing systems to provide secure solutions on the market and as well to have system readiness from eventual attacks. Almost all the software, that are launched in the market possess new and special features, but at the same time they carry with them several weaknesses that can be harmful. The above-mentioned problems cannot be fixed quickly, considering the complexity of such systems, where changes within the system structure need to be verified if it can affect some other part of the system. My paper based on the literature review and data collected from companies, will present a new framework for risk assessment maturity level.

Introduction

Once security and risk requirements have been identified and risk management decisions have been taken, appropriate controls must be selected and implemented to ensure that risks are reduced to an acceptable level. Controls are defined, as processes that helps detect, correct and prevent cases of unauthorized attacks or attempts at the organization's information [1]. The controls can be selected or can be designed to meet the specific needs of each case. The selection of security controls is dependent on organizational decisions based on the criteria for risk acceptance, risk treatment options, the overall risk management approach to be applied to the organization and should also be subject to all legislation and relevant national and international rules. Controls can be considered as guiding principles for information management security and are applicable to most organizations. Risk assessment, however, assumes risk identification; without it is difficult for organizations to measure the level of risk and as such its exposure would undermine the organization's security. Organizations that have an undefined status in relation to risk measurement are potentially vulnerable to possible attacks. Risk assessment is usually directly linked to the metrics. Risk assessment allows an organization to "recognize itself" about their risk exposures [2]. For this reason, the first step to determining the necessary protection measures for the information system is the process of risk identification. Risk identification is an important step in risk management, to determine what could cause a potential loss, and to gain insight into how and why the loss might happen. Thus, if a corporation expects to perform risk assessment successfully, finding the appropriate threat and vulnerability pair of each set is a crucial step. This paper will address a problem-solving model for managing the security risk assessment and will outline a framework that will address risk assessment. The structure of the article will start from the literature review, while continuing to present the structure of the proposed framework, which will ultimately present a prototype of the proposed framework. The article will be finalized with conclusions and future work to be done.

Literature Review

The purpose of the following literature review is to assess the state of the art in the field of information security respectively the risk assessment. I am studying the contributions of existing literature by gathering papers written so far and as well the case studies that have been implemented in relation to the research subject. Through the literature review I will identify gaps that have been studied in the field of information security and risk assessment. The identified gaps would provide me a better approach for my proposal and as well will direct my research and methodology. According to the literature the information systems security has evolved on three waves [3]. The first wave of evolvement is known as technical wave, which has been oriented on the technical approach to the information systems security. The second wave is known as the management wave which came as a result of the involvement of the management by understanding the importance of information and the need to protect those information's, while the third wave is the institutional approach where the code of conduct and the best practices have been adopted. The third wave according to [3] underlines that information security as a process that should be included in all the daily processes of employees and we must develop it as an information security culture across the organization. Organization are facing with more security risks, including organizations and technical

risks when there are digitalizing their services [4], [5]. Information security should be a concern for all computer users as nobody can avoid online attacks. That underlines, why it is important to know about the latest online security related issues. Information security is an important and decisive from protection of organizational assets' view. Information is an asset and as any other asset of an organization it has its value. According to SANS report and Manuel Finch [6], [7] information security refers to the processes and methodologies which are designed and implemented to protect print, electronic, or any other form of confidential, private and sensitive information or data from unauthorized access, use, misuse, disclosure, destruction, modification, or disruption. The existing and new businesses must face with the fact that information security risks may have a negative impact on the process of business continuity, public image, relationship between organizations, financial loss, affect relationships with clients, partners, and may create problems with legal authorities in case of discrepancies with the law. Information, support processes, systems and networks are important business assets. Defining, achieving, maintaining, and improving the security of information is essential to maintain a competitive edge, cash flow, profitability, legal compliance and trade image. Organizations, systems and networks face security threats from a wide range of areas including computer fraud, espionage, sabotage, vandalism, fires or flooding. Causes of malwares such as malicious code of programming, computer piracy and Denial of Service attacks (DoS) have become more common, more ambitious and more sophisticated [8]. According to the literature the information systems security has evolved on three waves [3]. The first wave of evolvement is known as technical wave, which has been oriented on the technical approach to the information systems security. The second wave is known as the management wave which came as a result of the involvement of the management by understanding the importance of information and the need to protect those information's, while the third wave is the institutional approach where the code of conduct and the best practices have been adopted. The third wave according to Solms [3] underlines that information security as a process that should be included in all the daily processes of employees and we must develop it as an information security culture across the organization. Organization are facing with more security risks, including organizations and technical risks when there are digitalizing their services [4], [5].

One of the biggest concerns for any organization is to achieve their goals by having their data protected. The information security is important for both types of public and private sector organizations, and it serves to protect critical infrastructures. In both sectors, information security will function as an incentive to avoid or reduce the risks involved. The relation between public and private networks and sharing of information sources increases the difficulty of achieving access control. The tendency to use distributed IT systems has also weakened the effectiveness of centralized and specialized control. Many information systems have not been designed to be safe. Security that can be achieved by technical means is limited and should be supported by appropriate management and procedures. Identifying which security control should be active requires careful planning and attention in every detail. Information security management requires participation by all employees in the organization. Participation of shareholders, suppliers, third parties, customers or other external parties may be required. Advice from specialists or organizations outside the main organization has often proved to be necessary.

My main research goal is to provide a new framework for security risk assessment which support the enterprises to analyze their information security maturity level. The framework will be based on the ISO 27001 and utilize the related standards as well. This risk management framework will be developed on a web-based software application and validated by companies from banks, IT and insurance companies in Kosovo.

ISMS: Information Security Mgmnt System

Diversity of opinions and factors influencing the process of IT adaption to information security needs is emphasized in many papers [9]. The literature has identified several factors affecting this process and most of them have listed factors such as senior management, government, IT consultants, organiz-

ational behavior, and so on [10]. Organizations are often affected by the models and standards that are implemented on information security within the same industry, but however not all the models and standards are implemented on the same way. For small organizations that operate with small staff and which distribute information with key staff only, the implementation of information security does not seem to be a necessary option. However, companies where information is distributed to more people simultaneously, it is impossible to manage them without a proper system, thus, presenting the problem of data vulnerability. The third group of organizations is on where the main product is information [11]. Organizations have different approaches when deciding to implement an information security system. Some organizations see, information security systems as a competitive edge in the market that can provide them with a greater credibility in their client relationship, as well as an increase of credibility in their organization and products. Another group of organizations implement information security systems only when they see that their competitors are operating in the same way. The views create cultural diversity within organizations of the same industry and no doubt enables them to improve.

Research Method

My research method will combine qualitative and quantitative methods according to Kaplan [12]. The primary research method for this study is literature review and interview with companies using questionnaire. This study will first review various types of risk assessment methods and then as the second phase I will compare some maturity model's frameworks and their characteristics. In the third phase, I will prepare a questionnaire and will conduct interview with the persons responsible for Information Security at the IT, Banking and Insurance Company sector such as Chief Information Security Office, Data Protection Officer, Information security and assurance, Risk Management Officer etc. depends on the company structure and people responsible. Interviews will be conducted face to face and the structure of questions will be, multiple choices and as well open questions face-to-face, one-on-one interviews with key personnel within an organization are the primary method for data-gathering in case studies and are essential sources of information [13], [14]. Through the interviews I will collect data and will analyze them with a statistical software such as SPSS, Stata or R-Statistics to get as more accurate results. Based on this understanding, a method will be developed to compare frameworks according to their levels of maturity, issues that treat, method of implementation and their compliance. Last stage of this study, existing models and the comparison will be identified based on a comprehensive review of current industry practices and academic researches. Finally, once the maturity models are identified and the comparison between them has been made, I will propose a new approach on maturity model be outlined.

My approach will be followed by developing the software prototype which will be based on my approach. The software will be a web-based application developed on PHP programming language and the database will be based on MySQL. The web-based application will be optimized for use on every device from Desktop PC to Tablets and Smartphone with the technology of auto responsive content. This means that depends on the resolution and the screen of the device, the software will be automatically optimized. This application will be user friendly and easy to navigate through it and with less memory and internet consuming which we will solve by implementing the backend-oriented layout using the HTML5 and CSS3 mostly for design and very few images. On completion of the questions from the companies and organization this system will have the opportunity to export the report generated with the recommendations. The prototype will be test before the organizations can use and as well it will have a period as beta version during which time we will identify any bug or possible improvements.

Research Scope and Question

My research is based on four maturity model frameworks respectively, ISM3 (Information Security Management Maturity Model), SSE-CMM (System Security Engineering Capability Maturity Model), and NIST Maturity Model. I determine the gaps from the above-mentioned maturity models and prepare my maturity model according to the previous investigation. There are several IT risk management

approaches, good example and frameworks reported in the literature, an examination of which reveals some development in thought processes regarding the most effective risk management method. It is necessary for risk professionals and auditors to have a maturity model through which they can check if their risk tools are meeting the expectations and producing results. The risk management standard that we use in organization is not important, if it helps us to make better strategies for risk management. Many risk management programs have a built in Risk Maturity Models themselves, which can be broken down into many other sections focusing on core attributes [15]. A recent third-party study on Risk Management solutions proved that enterprise risk management maturity as calculated by Risk Maturity Models added 25% to a business organization bottom line value [16]. Siponen [17] used a term "software maturity criteria", by which he proposed and explained that existing maturity standards must lead the way toward the managing the information security in organizations. Poeppelbuss [18], explained maturity models as a conceptual model that describes a way of how the organization will outline the logical and desired evolution toward maturity, while the maturity from (Bruin and Rosemann, 2005) is described as an evaluation measure for the organizations capacity to follow a certain discipline. Another opinion regarding the maturity has been provided by Mettler [19] who describes the maturity as an progress of evolution on demonstrating the ability to accomplish an targeted activity from beginning to the final stage. Recently, there is an increased interest for the maturity models in the research community and in the practice [18], [19]. The most popular maturity model is Software Engineering Institute's (SEI) Capability Maturity Model (CMM) for software development and the successor Capability Maturity Model Integration (CMMI) [18]. Until now several new maturity models have been developed for different sectors and industries: IT/business alignment [20], [21]; business process management [22], business intelligence [23]; project management [24]; information lifecycle management [25] digital government [26]; inter-organizational systems adoption [27] and enterprise resource planning systems use [28]. According to the literature [29], there is a gap between the implementation of the information security standard and business sector needs and objectives. In this regard I will address the following research questions:

Research question: Is it possible to measure the maturity of information security based on ISO 27001 within a company through a semi-automated risk assessment system?

Answering this question, I analyzed literature that describes the digital maturity models in the context of information security. In most cases, automated processes for risk assessment have been used only by audit firms. After a detailed analysis of literature and cases I will list the most appropriate approaches that efficient to identify findings within organizations' systems.

Standards related to ISMS

ISO 27000

According to ISO (International Organization for Standardization) ISO 27000:2013 refers to standard family which provides organizations with a standard for information security management and general structure for the management system. This standard is created by a wide variety of organizations and compiled by the International Organization for Standardization [30], [31]. ISO 27001:2013 covers the establishment, implementation, maintenance and continual improvement of an information security management system. It also has requirements to assess and treat information security risks. All the requirements set in the ISO 27001:2013 are generic and intended to be applicable to all organizations, regardless of the size or nature [32], [33]. ISO 27001 is one of the most widely adapted information security management frameworks [34], [35]. It is a framework for establishing an effective information security management system (ISMS). My research will be based on this standard because this framework is widely accepted in the field of information security. It has a top-down approach and it is based on risks, which means that the framework is technology independent. One of the first requirements during the implementation of ISO27001 is the definition of risk assessment within the organization. According to

standard requirements, the risk assessment methodology should be based on business, information security as well as other legal and regulatory requirements that enable accurate identification of the level of risk. The ISO 27001 documentation also describes the need for the organization to be able to identify assets, risks and identify system weaknesses. The maturity model in ISO 27001 can be described on several points such as by comparing and measuring the benefits with previous projects implements, circumstances that can gather different goals, the model for definition the priorities etc. Hence, it helps us to use the maturity models as a comparative tool to understand what we are expecting from the organization.

As I have already mentioned, risk assessment is a process that can be considered as an independent process from technology implementation within an organization. Furthermore, the need for its centralization is very important, being a process that also helps to identify problems, risks and threats possible in organizations. As such, the security of information is considered a process that contains many activities within it, and this has driven the need for information security to be integrated into maturity models. Improving the security of information within the organization affects many other processes and may also affect changes in the business strategy, so it is known as an important and long-lasting process that cannot be changed and applied over the night. Management focuses on proving the information security strength of the organization by implementing information security into the organizational culture, certification, and continuous measurement and monitoring of risk assessment processes.

CMMI

The Capability Maturity Model Integration (CMMI) defined by the Software Engineering Institute (SEI) of Carnegie Mellon University is gaining in importance in Europe. It is an effective tool that helps to improve the effectiveness and efficiency of development organizations [36]. One of the strengths of CMMI is its specialization in product development. This makes it possible to focus on specific aspects in a much more precise and in-depth manner than generalist models such as ISO 9000. For each proposed practice, CMMI provides one to two pages of bullet points and descriptions that can serve as a guide to improvement [37].

Compared to other specialized process models for development organizations, CMMI has the advantage of bringing together different views of the organization. It addresses project management, development, organizational support, process improvement, and management tasks in a common model.

NIST

National Institute of Standards and Technology (NIST) maturity model focuses on documentation of procedures [38], [39]. This NIST framework is defined on five maturity levels such as: Policy, Procedure, Implementation, Testing and Integration in which the information security is considered as a risk that is managed through the enterprise risk management process. According to this we have identified the NIST framework as a risk-based framework [40]. The focus area of NIST maturity model is to check the level of documentation [41], [42].

ISM maturity model ISM3

ISM3 represents one of the standards from the information security area whose main goal apart from achieving the admissible level of security is achieving the business goals. ISM3 is a process-oriented approach and according to this management activities must follow different categories of the process such as: Risk assessment which discovers the treats, attacks and vulnerabilities. The ISM3 was introduced to prevent and mitigate attacks, error and accidents that may risk security [43]–[45].

Software for Risk Assessment

There are various software applications for different models, techniques and different methods of risk analysis. These software's use methods and techniques such as questionnaires, checklists, passive assessment, active evaluation in various versions to obtain appropriate risk analysis information. Before we decide which application to use, we need to define the testing process we want to apply for

example if we are dealing with the overall assessment of the company, we can use applications that have the form of the questionnaire or if we want to test any organization software then we can we use apps that make active evaluation. In this case, active assessment means using an application to test the organization's software stability by doing attacks in various forms such as password attack, database attack, phishing attacks, and so on. In some cases, applications are built on the functional structure of the models.

FAIR software is the "quantitative risk engine" for the FAIR model, which's main goal is to find the source of risk. This software achieves this by taking measurements of risk factors and applying sophisticated mathematical principles to find the risk. FAIR provides a centralized "warehouse" of analysis to have a general overview of the risks, an overview of the accumulated risk, a simple view of risk comparison for their prioritization, a centralized asset database, potential risks, tables losses, users, graphs, etc. Like the FAIR model, this software is quite complete as it is a combination of some models and is complemented by some other software. This software delivers results of risk factors, why it happened and how it happened, but focuses on extracting accurate quantitative results.

Octave software is used to identify and assess the risk of information security. They try to help organizations set up quality risk assessment criteria that describe the level of company tolerance to operational risk, identify assets that are important to the organization, identify threats and weaknesses to these assets, assess potential damages to the organization if the risk is realized etc. For the OCTAVE model there are several software generations, created by CERT in a way that we have different access to information assets and the elasticity in their use. The latest models like OCTAVE Allegro have not been created to replace the pre-models, but to create selection varieties. However, each version of OCTAVE has wide applicability and users of these methods can choose the approach that best suits their security risk assessment information needs.

CRAMM, as a matrix model, depends heavily on supporting software to provide full support. This software serves to analyze and manage quality risk. This tool was built by the UK government to provide a method for reviewing security information systems. The CRAMM Manager can be used to justify costs in the security of information systems and networks and testing of standards compliance for the certification process.

Risk Assessment Maturity Framework

Answering to my research question, I will provide my proposed framework on Risk Assessment Maturity. The main goal of the system is providing a framework for internal auditors, information security experts and other interested parties and stakeholdersto perform risks assessment in a company.

My proposed framework is based on reviewing some of the current frameworks such as Fair, Octave and Cramm we have seen that most of the frameworks are quite complicated to be used, with many functions and complex processes. The framework is targeting end-users, so the main requirements are: user-friendliness, easy to be used, and adaptable to develop any risk assessment questionaire. The application is made up of several blocks that represent the respective functions as well as are interconnected with other parts of the system. In the dashboard of the system, we have presented various statistics that show, the number of companies that have carried out the risk assessment, the number of questions, how many questionnaires have been conducted and how many questions have been answered. Also, other statistics are visualized on the dashboard, such as the most frequent answers, the most frequent security issues from all questionnaires and so on

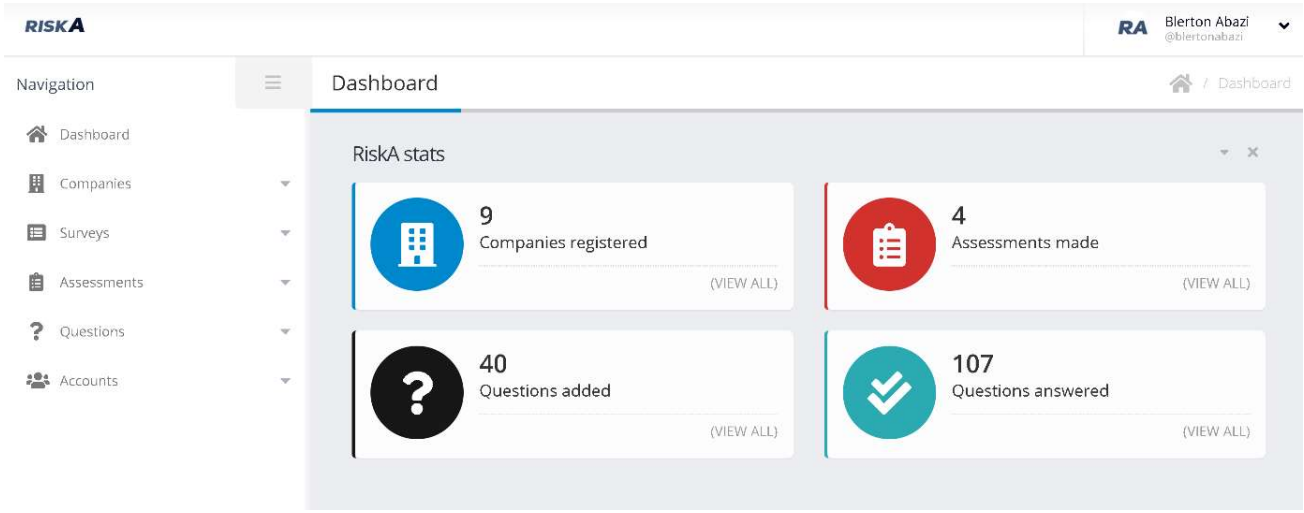

Figure 1 The system dashboard

The application also has a navigation menu on the left that helps us to overall manage the system. In the navigation menu we have 6 sections as follows:

- *Dashboard:*presents visualizes general data and statistics
- *Companies:* This section helps us to obtain general data for companies that will be subject to the questionnaire. In this section we have two subsections, respectively the option to register new company and the current list of the companies that are already on the system
- *Surveys:* – This is the main part of application because through this section we manage with questionnaires. In this section, we can add new questions from the database, categorize questions, or even change the type of the questions.
- *Assessment:*- In this section we can see the list of assessments we have accomplished so far. Particularly in this section is that we can make a comparison between some assessments. For example, if Company X has conducted the Assessment in 2017 and 2018, then through the *Compare Assessment* option we can see the progress that the company has made in certain sections.
- *Questions:* through this section we can add new questions, modify the existing ones, or even change the form of the question.
- *Accounts:*- is the ultimate part that enables us to administer the system or create new users by setting the level of use. For the moment we have two types of users, respectively administrator and user simple.

Looking at different models of software applications that make risk assessment, based on different techniques and methods, we have found it reasonable to create our model as well. To build this application we used the questionnaire technique. The application is built on web technology, as it provides easy and fast access from various devices and wherever there is Internet access. The technology used for the user-interacting look is developed with HTML, designed and stylized with CSS and Bootstrap, animations and JavaScript behaviors. To have dynamic content, to display the questionnaire etc., in the background for data manipulation is used PHP and data storage is used by the MySQL database. The software is concepted in such a way that only authorized persons with specific privileges can access to the system, and every use and manipulation of the system is recorded on a log sheet behind the system. Once one of these people accesses the system, he/she can create different types of questionnaires based on the assessment that he/she want to make.

So, we can create any questionnaire while and each questionnaire that contains sections or subcategories. Sections should contain questions related to a particular topic. Questions can have up to 5 responses to be predetermined, and each having its

own value. Once we have filled this data, we can create different versions of the questionnaires and provide manipulation with sections belonging to questionnaires, as a section may be in a different questionnaire, and a questionnaire may have many sections. If the creation of questionnaires has been completed, we can continue with the registration of companies that are subject to the risk assessment process. Only simple, informative information about the company is required, and we may be willing to continue with the next steps. Each question may have different types of response tailored to each case, as there is a possibility to change 5 response levels as needed. Whenever we want to add a new question and do not have the desired options, a new set of options can be added and used in the new question. A set contains more than 5 options, all with the option of adjusting as needed. Although the answers to the questions are presented with a rating of 5 options ranging from 1 to 5 points, this does not mean that the analysis is quantitatively. In the application there is the possibility that numbers can easily be replaced by word or sign and have the same meaning. These answers may represent frequency, method, concrete response to Yes and NO, etc. After answering all questions in all sections, we can go to the next page, so all the answers are stored based on the data. From the answers we have provided, the result will be calculated once for each question, and it will show as a result an average response per section and a general average. For each section we will have a result and a recommendation based on the level of responses.

Conclusion

Risk assessment applies primarily because it creates a safe working environment, protects the organization's employees, protects assets, protects the company's reputation, and so on. Risk forecasting is vital to maintaining the company's assets and ensuring good performance in the company. It enables us to recognize dangerous factors, alert workers with the situation, create a standard of management with safe practices and legal requirements, reduce unsatisfactory events, etc. A serious threat to major organizations is third-party software, which has created bad examples in the past. When the organization makes collaborations with third parties, the chances are that resources should be provided to this party for cooperation opportunities. Third parties get access to the organization's assets and if an incident occurs to these partners, the organization's resources are immediately compromised. Third-party use is always tricky, as we cannot apply a risk assessment of the company and their resources before cooperating with them. Another important issue is the risk management which relies on two strong pillars, one is risk identification and analysis, the other is risk response and monitoring. The most common way of dealing with risk is the reduction, which does not mean that it should always be used, but often gives a good result. Reduction does not mean that we have made a permanent solution, continuous monitoring is indispensable. The transfer of risk through insurance or contract is a growing trend and is expected to have a major impact on the future of risk management. A risk would bring enormous damage to the company, and management with it is too costly or the organization does not have enough resources to manage that risk, then the risk transfer of an insurance company would be a matter of choice.

References

[1] ISACA, *CISA review manual 2007*. 2006.
[2] M. Talabis and J. Martin, *Information Security Risk Assessment: Risk Assessment*. 2012.
[3] B. von Solms, Information Security — The Third Wave? *Comp. Secur.*, vol. 19, no. 7, pp. 615–620, 2000.
[4] M. Amberg, R. Markov, and S. Okujava, “A Framework for Valuing the Economic Profitability of E-Government,” in *International Conference on E-Government (ICEG)*, 2005.
[5] A. Brown, “IS Evaluation in Practice.,” *Electron. J. Inf. Syst. Eval.*, vol. 8, no. 3, 2005.
[6] SANS, “Information Security Resources,” 2008. [Online]. Available: https://www.sans.org/information-security/.
[7] B. Maule-Ffinch, “Key trends in information security,” *Netw. Secur.*, vol. 2015, no. 11, pp. 18–20, 2015.
[8] McAfee, “2017 Threats Predictions,” no. November 2016, p. 39, 2017.
[9] J. Businge, A. Serebrenik, and M. van den Brand,

"An Empirical Study of the Evolution of Eclipse Third-party Plug-ins," in *Proceedings of the Joint ERCIM Workshop on Software Evolution (EVOL) and International Workshop on Principles of Software Evolution (IWPSE)*, 2010, pp. 63–72.

[10] A. Joshi, L. Bollen, H. Hassink, S. De Haes, and W. Van Grembergen, "Explaining IT governance disclosure through the constructs of IT governance maturity and IT strategic role," *Inf. Manag.*, no. January, pp. 0–1, 2017.

[11] D. T. Burgeois, *Information Systems for Business and Beyond*. 2014.

[12] B. Kaplan, D. Duchon, and A. C. Study, "Combining Qualitative and Quantitative Information Systems," vol. 12, no. 4, pp. 571–586, 1988.

[13] S. Kvale, "The Interview Situation," *Interviews: An Introduction to Qualitative Research Interviewing*. pp. 124–159, 1996.

[14] R. K. Yin, *Case Study Research: Design and Methods*, vol. 5, no. 5. 2009.

[15] Sophia Wright, "How Can Risk Maturity Model Benefit Your Risk Management," 2014. [Online]. Available: https://www.riskmethods.net/en/blog/How-Can-Risk-Maturity-Model-Benefit-Your-Risk-Management/112.

[16] S. M. Sarah Beals, Carol Fox, "Why a mature ERM effort is woth the investement," *Exec. Rep.*, p. 5.

[17] M. Siponen, "Towards maturity of information security maturity criteria: six lessons learned from software maturity criteria," *Inf. Manag. Comput. Secur.*, vol. 10, no. 5, pp. 210–224, 2002.

[18] J. Poeppelbuss, B. Niehaves, A. Simons, and J. Becker, "Maturity Models in Information Systems Research: Literature Search and Analysis," *Commun. Assoc. Inf. Syst.*, vol. 29, no. 1, pp. 506–532, 2011.

[19] T. Mettler, "A Design Science Research Perspective on Maturity Models in Information Systems," Institute of Information Management, Universtiy of St. Gallen, St. Gallen, Oct. 2009.

[20] J. N. Luftman, "Assessing Strategic Alignment Maturity," in *Competing in the Information Age: Align in the Sand: Second Edition*, 2003.

[21] M. Khaiata and I. A. Zualkernan, "A simple instrument to measure IT-Business alignment maturity," *Inf. Syst. Manag.*, vol. 26, no. 2, pp. 138–152, 2009.

[22] T. De Bruin and M. Rosemann, "Towards a Business Process Management Maturity Model," in *ECIS 2005 Proceedings of the Thirteenth European Conference on Information Systems*, 2005, pp. 1–12.

[23] P. Hewlett, "The HP Business Intelligence Maturity Model," 2007. [Online]. Available: http://h20195.www2.hp.com/v2/GetDocument.aspx?docname=4AA1-5467ENW&cc=us&lc=en.

[24] J. Kent Crawford, "The project management maturity model," *Inf. Syst. Manag.*, vol. 23, no. 4, pp. 50–58, 2006.

[25] Sun Microsystems, "Information lifecycle management maturity model," no. April, pp. 1–8, 2005.

[26] P. Gottschalk, "Maturity levels for interoperability in digital government," *Gov. Inf. Q.*, vol. 26, no. 1, pp. 75–81, 2009.

[27] M. Ali, S. Kurnia, and R. B. Johnston, "Understanding the Progressive Nature of Inter-Organizational Systems (IOS) Adoption," 2011.

[28] C. P. Holland and B. Light, "A Stage Maturity Model for Enterprise Resource Planning Systems Use," *SIGMIS Database*, vol. 32, no. 2, pp. 34–45, Apr. 2001.

[29] B. Von Solms and R. Von Solms, "From information security to...business security?," *Comput. Secur.*, vol. 24, no. 4, pp. 271–273, 2005.

[30] International Organization for Standardization, "ISO 27000 Directory," 2014. [Online]. Available: http://www.27000.org/background.htm.

[31] G. Disterer, "ISO/IEC 27000, 27001 and 27002 for Information Security Management," *J. Inf. Secur.*, vol. 04, no. 02, pp. 92–100, 2013.

[32] International Organization for Standardization, "ISO. 2013. ISO/IEC 27001 – Information security management," 2014. [Online]. Available: http://www.iso.org/iso/home/standards/management-standards/iso27001.htm.

[33] B. Shojaie, H. Federrath, and I. Saberi, "Evaluating the effectiveness of ISO 27001:2013 based on annex A," *9th Int. Work. Front. Availability, Reliab. Secur. (FARES 2014)*, no. Fares, pp. 259–264, 2014.

[34] S. Wright, "Measuring the Effectiveness of Security using ISO 27001," *Inf. Warf. Site*, pp. 1–15, 2006.

[35] K. Beckers, S. Faßbender, M. Heisel, J.-C. Küster, and H. Schmidt, "Supporting the Development and Documentation of ISO 27001 Information Security Management Systems through Security Requirements Engineering Approaches," no. 256980, pp. 14–21, 2012.

[36] R. S. Tapia, M. Daneva, P. Van Eck, and R. Wieringa, "Towards a business-IT aligned maturity model for collaborative networked organizations," *Enterp. Distrib. Object Comput. Conf. Work.*, vol. 12, pp. 276–287, 2008.

[37] L. Greiner, "Capabilit Maturity Model Integration (CMMI) Definition and Solutions," no. Cmmi, pp. 1–10, 2018.
[38] X. Y. Ge, Y. Q. Yuan, L. L. Lu, G. Xiao-yan, Y. Yu-qing, and L. Li-lei, "An information security maturity evaluation mode," *Procedia Eng.*, vol. 24, pp. 335–339, 2011.
[39] A. Johnson, "Guide for Security-Focused Configuration Management of Information Systems," *Nist*, no. August, pp. 1–88, 2011.
[40] NIST, "Framework for Improving Critical Infrastructure Cybersecurity," *Natl. Inst. Stand. Technol.*, pp. 1–41, 2014.
[41] D. a. Chapin and S. Akridge, "How Can Security Be Measured?," *Inf. Syst. Control J.*, vol. 2, pp. 43–47, 2005.
[42] S. Woodhouse, "An ISMS (Im)-Maturity Capability Model," in *2008 IEEE 8th International Conference on Computer and Information Technology Workshops*, 2008, pp. 242–247.
[43] V. Aceituno, "Information Security Management Maturity Model," 2007. [Online]. Available: www.ism3.com.
[44] Open Group, "Open Group Standard Open Information Security Management Maturity Model," *ISM3 Consortium*. Van Haren Publishing, 2011.
[45] B. Stevanovi, "Maturity Models in Information Security," *Int. J. Inf. Commun. Technol. Res.*, vol. 1, no. 2, pp. 44–47, 2011.

Efficient Energy Management System and Optimization of Resources at a Furniture Company

[1]Adrienn Koncz, [2]Attila Gludovátz
[1,2] Institute of Information Science and Economics, University of Sopron
eMails: koncz.adrienn@uni-sopron.hu; gludovatz.attila@uni-sopron.hu

Abstract

The Industrial Internet of Things is connected to the 4th industrial revolution, which is currently a very actual topic in the industry and the information technology too. The 4th industrial revolution have influence on the production processes with new tools and new technologies. Industrial companies continuously develop and optimize all manufacturing, production and market processes. Data collection and analysis are essential for the process controlling. In general, production processes are the main issue for the industrial companies, because their main goal is to maximize the productivity and profit, or to minimize the loss.

At an IoT (Internet of Things) company, the most important tasks are placing data collection equipment and implementing such system which can measure and collect production data. There are some data collectors which measurements are not so closely connected to production data, though, these are very important equipment and without them the production processes cannot work. An IoT company needs reliable measuring instruments, which can measure the total and partial electricity consumption (kWh), amperage (A), voltage (V), water and generated air consumption. The smart measuring instruments provide the data to the center with the help of specially built internal network. Data are displayed in real time in the building management (SCADA Supervisory Control and Data Acquisition) system [1]. The management can interpret the received data through the monitoring and supervisory system, and they can make analyses and reports from the data.

Introduction

In this study, we demonstrate the energy management system framework and its operation and structure at a furniture company. We created the system for data collecting. These historical and real-time datasets are the basis of the analysis and it helps us to get the information of the machines, products, energy and resources consumption [2]. We can make estimates for the future events with forecasting methods. In addition, we can focus on the occurrences of critical events with defining the limit for risky parameters. Our goal is to prevent the stoppage of the production or the excessive use of the energy and other resources.

At the beginning of our work, we focused on the data collecting process. The company's supervisory system provided us the necessary energy usage data of the selected manufacturing machines. We chose different machines, from the aspect of utilization of the energy (high, moderate, standard and low). We made the measurements with the data-collector sensors [3]. We built the communication network between the machines, sensors and the data-centres. In this way, we could keep under control the complete process in real-time. The energy usage of the machines could be constant, but there is no consideration of the various manufacturing at the company.

The company produces circa 1500 different types of products. In addition, there is an upper limit of the energy usage a day, on average, that the company should comply with it. So then, we must collect and select the data, which are connected to the production process. At the end of this process, we joined the total sum of the fact data through the dimension data (the common parameters: the rounded timestamps and machines' identifiers). We made the merging progress in 10-minutes periods. We can analyse the time series of the various data with help of a business intelligence software.

Beside energy consumption, the company wanted to know other consumption data of the firms, that is why we made some developments at the company's plants a few months ago. We started to measure the water consumption of the firms. We installed sensors to the compressors to measure their generated air quantity beside the compressors' electricity consumption.

In the following section, we introduce the literature review about industrial internet of things, the industrial energy management subject and our researching methods. After this, we demonstrate the operation and structure of the energy management system framework. We represent our energy consumption results with diagrams. After the energy management phase, we demonstrate our research and results in connection with the management of resources. In that section, we presented the water consumption data and our estimates related to this subject, after that, we will introduce the structure of the compressor housing and the generated air quantity / the compressors' electricity consumption (rate) data. The final section contains the conclusion of the study.

Literature review

Many definitions exist for Industry 4.0, but if we would like to express and explain to the point, we can say that "*the industrial internet is an internet of things, machines, computers and people enabling intelligent industrial operations using advanced data analytics for transformational business outcomes, and it is redefining the landscape for business and individuals alike*" [3]. Another definition says that *"Industry 4.0 is a collective term for technologies and concepts of value chain organization. Within the modular structured Smart Factories of Industry 4.0, CPS[3] monitor physical processes, create a virtual copy of the physical world and make decentralized decisions. Over the IoT, CPS communicate and cooperate with each other and humans in real time. Via the IoS [Internet of Services], both internal and cross- organizational services are offered and utilized by participants of the value chain"* [5].

Industry 4.0 is existing more main areas and subjects: (1) Artificial Intelligence, (2) Cloud Computing, (3) Additive manufacturing, (4) Horizontal and vertical Integration, (5) Big Data, (6) Machine Learning algorithms, (7) Autonomous robots and

[3] Cyber-Physical System [10]

robot programing, (8) Indoor positioning system, (9) Collaborating and sensitive robots, (10) Digital twin, (11) Interactive production design simulation (12) Augmented and Virtual reality (AR and VR), (13) MES[4]–ERP[5] system connection, (14) IoT production management, (15) Digitized energy management [15] [16] [17] [18]. We are dealing with the "Digitized energy management" theme, within that the optimal use of resources (raw materials, electricity, water, generated air etc.). We studied and examined the scientific literature background of this subject.

It was considered necessary to launch a project that is linked to Industry 4.0. We set up a research methodology, which is based on the following parts and steps: (1) The first and most important thing is that the company's management makes a strategic decision, which contains the next answers for these questions: "*In what do we want better than others?*", or "*What process do we want develop and optimize?"*; (2) We have to get to know deeper the selected process in-depth with data collecting: (2a) To install the data collectors (sensors) to the equipment OR in a special way, with the help of a smart camera, which delivers the signals, which can be transformed to digital data, (2b) To make a connection between the sensors and centres, (2c) To store the incoming data into the database, (d) To gain the data from other source systems (optional step); (3) Raw data analysing with visualization applications: Discovering the problems and correcting them; (4) Data analysing with mathematical, statistical, time-series and data mining techniques; (5) Presenting the results to the decision makers; (6) Making decisions for optimization; (7) To operate the system and supervise, control, analyse the processes all time. We support these steps of our methodology with IT toolkits and techniques.

We work together with a multinational furniture company, where we have the opportunity to use our framework. In this paper, we will introduce our methodology in practise at this company and represent our results too.

Energy management system

In the last years we developed the company's ERP system and its energy management system. The company needed these developments, because they wanted to know more data about their production and manufacturing data and their energy consumption. That is why we created a new energy management system's framework. Before the installation of our framework, the company knew only one data in reference to the daily energy consumption. The production lines, machines and other units were equipped with smart measuring instruments, which measure the electricity consumption. The measuring instruments were installed into a new computer subnet at the company, for that they cannot cause failures to the other communication networks [6]. The data can be tracked and traced with the help of the applied supervisory system at the company, however there was no possibilities to drill down in the dataset at a deeper level [7]. Because of it, the measured parameters (for example: the machines' electricity consumption) were saved in an external database with the help of periodical incremental backups. In this database, we designed a universal data structure, in which any numerical data can be stored with their descriptive parameters. At the same time, we got to know the selected manufacturing machines' production data, which are coming from the company's ERP system 0[9].

The goal was that we can connect production data to the energy consumption data. The connection points were the fact data's timestamps and the identifiers of machines. At the beginning of the research, we defined a 10-minute time period. So, one energy data row generated in every 10 minutes to every examined machine, that process was an incremental savings. The frequencies of production data were different: there were more data row in a 10-minute period. To join these datasets, we had to make several summarizing (production quantities) and rounding (production and energy consumption rows' timestamps) transformations on the datasets. The joins of datasets were made by a business intelligence tool (with SQL scripts). In this way, the

[4] Manufacturing Execution Systems

[5] Enterprise Resource Planning

necessary joined datasets (production and energy consumption data) are available for monitoring and analyzing. The configuration of the system is well-manageable, there is no need high level IT knowledge. Since the installation of the system, the company's specialists did not only measure the electricity consumption, but the company's generated air and water consumptions.

One of our main goal was that we support the energy efficiency and the production process with IoT solutions and applications. Because of it, we created an expanded cyber-physical system at the company. Our other goal was that we add minimum 10 manufacturing machines and numerous supporting equipment into the operation of the framework. With this system, we can make better decisions, because, we can analyse the energy consumption data and the other relevant production data. So, we can get useful results from the data and thereby we can get answers to managers' optimization questions.

These questions can be classified into two major groups. The first group is related to production machines, their energy consumption and their production quantity databases. Based on these data, we can analyze the production and energy consumption parts jointly. The second group includes the other support equipment (compressors, ventilators, compressors etc.) which appear as energy consumers in the plants. We created the above presented framework that we can answer all of our questions and we can help the company's management to make more efficient decisions related to the production and energy consumptions subject [12].

Electricity consumption

In this phase, we introduce our results of analysis, that we made from the measured energy consumption data and production quantities. We made several reports, which are helpful for the management. At first, we present some diagrams and results about the relation of electricity consumptions and production outcomes of machines (produced pieces/kWh), after it we analyse the electricity consumptions of the other equipment.

In the Figure 1 we can see the electricity consumption of all machines. The dates can be seen on the horizontal axis, which defines an almost complete financial year (2017 September – 2018 March). The kWh energy consumptions are on the left vertical axis. The light grey columns show the useful energy consumption, the dark grey columns mark the useless energy consumptions. The numbers of created product pieces are on the right vertical axis, it's represented by the line series. In this case the useful energy consumption means when the machines were produced, and the useless energy consumption means when there was no production process. At the end of the period, we can see only dark grey columns, because the company changed the version of ERP system, and that is why, there were no production data for a short time. When the update of the new ERP system version was ready, the data of two plants could receive in a large integrated database. Earlier in each plant, there were two databases with the same structure and data tables. After the version update, the production data arrived again flawlessly.

We made the same analysis with the data of each machine (see Figure 2). As an example, we present the data of a shaving machine. The axes and the series are the same as the latest diagram. At the beginning of the financial year, we did not measure the energy consumption of this machine. At the end of the financial year, we did not get any production data because of the ERP system version changing.

We examine the weekly data of another machine, that we can see in Figure 3. During the examining, we identified that this shaving machine is an uneconomic machine. The colours of the columns' marks are the same as the first diagram (the dark grey columns mark the useless energy consumption and the light grey ones mark the useful energy consumption). As we can see, this diagram indicates more useless energy consumption quantity during the machine's manufacturing process. This extraordinary result had our attention, that something was wrong with the machine. We summarized the data in every 10 minutes, and we saw that there was more 10 minutes period when the machine consumed electricity, but it did not produce. Later the system

indicated the machine produced hundreds of pieces in 10 minutes, which is impossible (Figure 4). It happened because the operator (who controls the manufacturing machines) recorded the data manually at the end of the batch. Because of it, a lot of 10-minute consumption data seemed useless, and per batches only one 10-minutes period was the energy consumption useful.

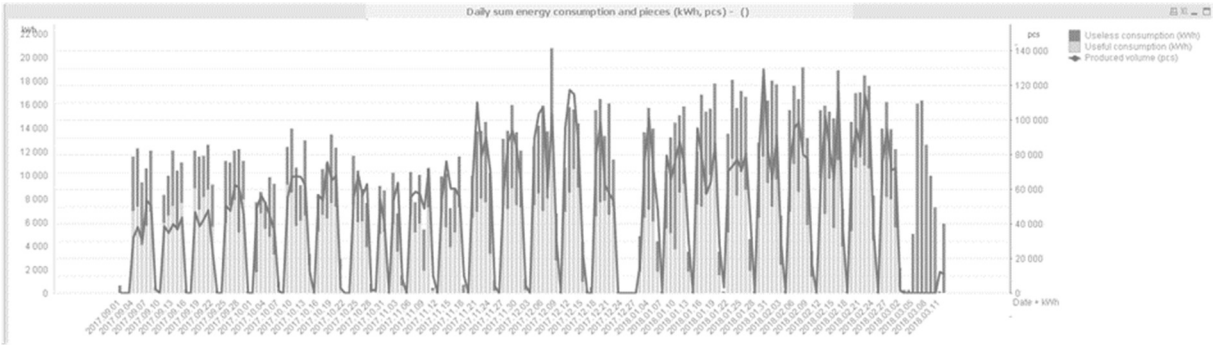

Figure 1. Electricity consumption of all machines
(Source: company's BI[6] system)

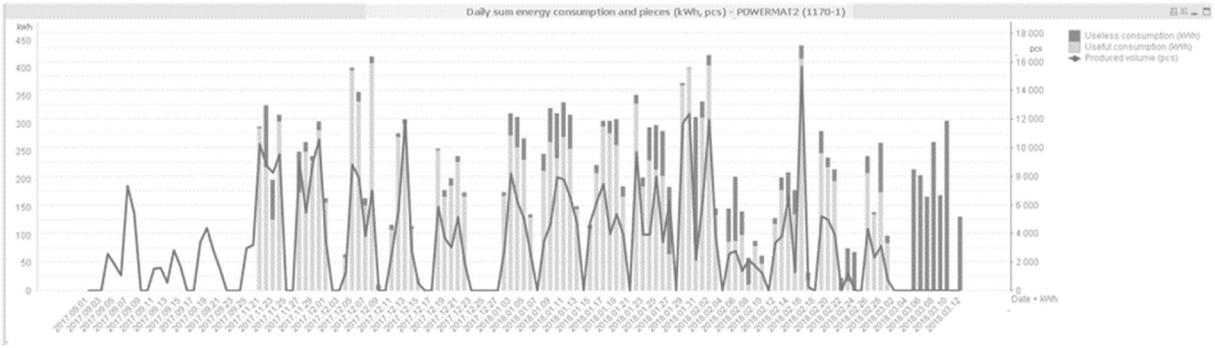

Figure 2. Electricity consumption of a shaving machine (Source: company's BI system)

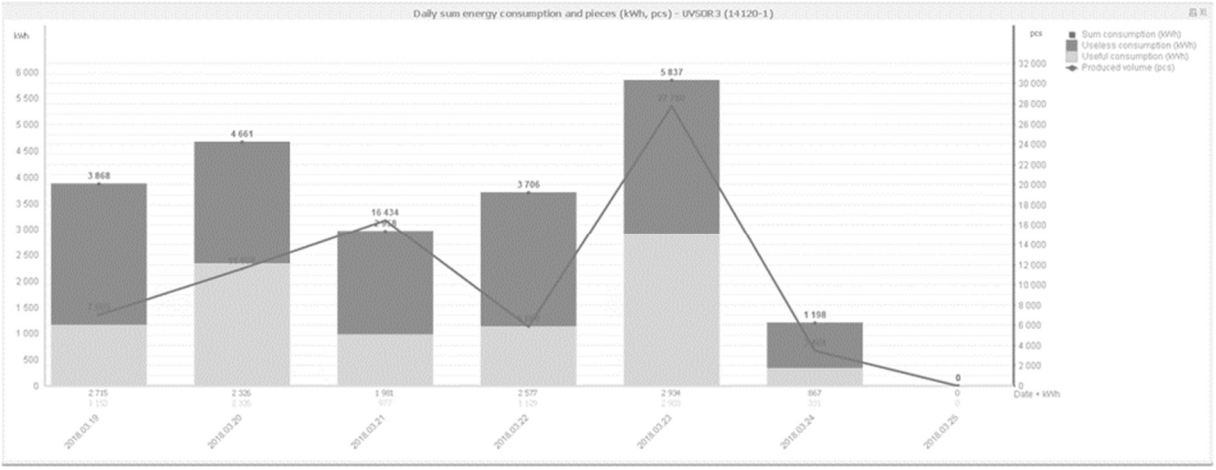

Figure 3. One week's data of a manufacturing machine (UVSOR3 - 19/3/2018 – 25/3/2018)
(Source: company's BI system)

[6] Business Intelligence

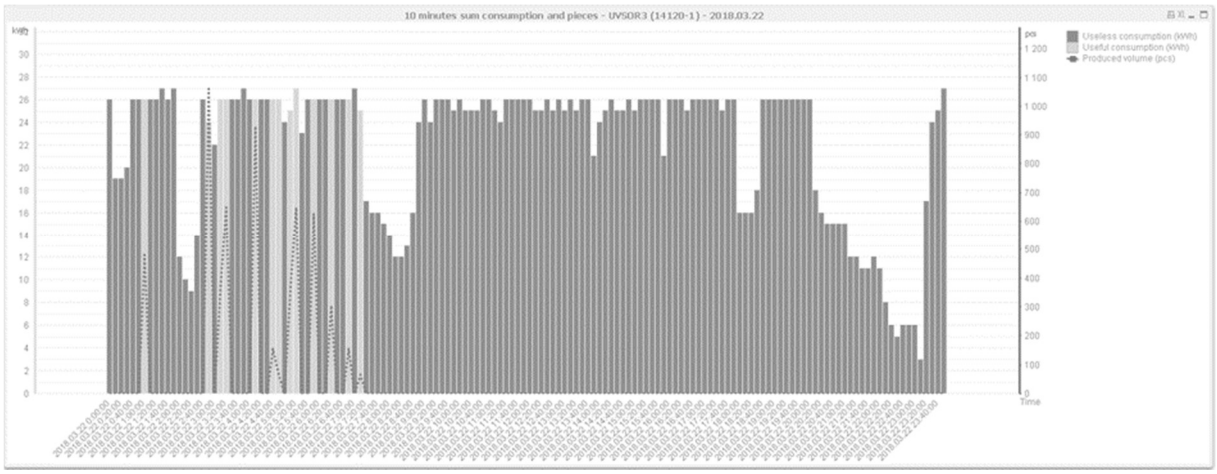

Figure 4. The daily performance of a selected machine (UVSOR3 - 22/3/2018) (Source: company's BI system)

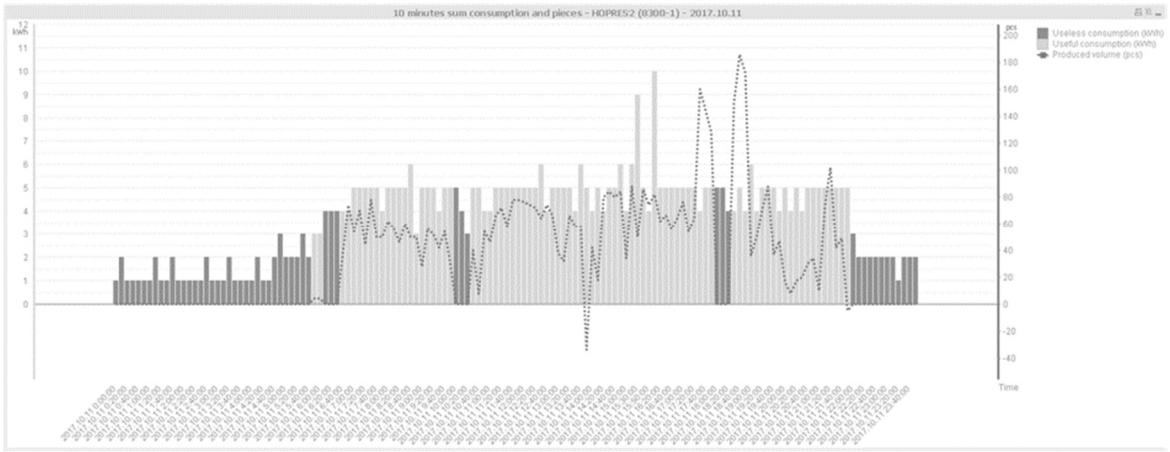

Figure 5. The daily performance of the manufacturing machines (heat press machine) (Source: company's BI system)

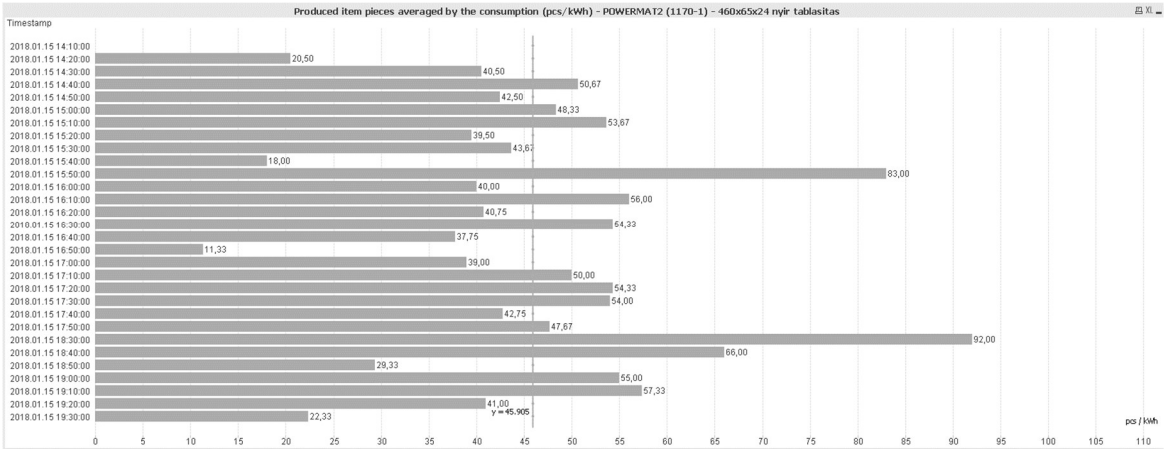

Figure 6. Production efficiency of a given product and machine. The horizontal axis shows the values of "Comparison 1. (number of pieces/kWh)". The vertical axis shows the dates when the given product has been produced. (Source: company's BI system)

Figure 5 shows a daily operation of the heat press machine. We can see that the machine was left in standby mode during the night, and it consumed 1-2 kWh in every 10 minutes. At the beginning of the morning shift some workpiece were made with the machine. During this shift the workers continuously produced, but they had a break (from 10:00 to 10:20) when the machine doesn't produce any product. We can see that they did not put the machine in standby mode to consume less, they just left it during its operation. It appears an interesting thing that the line chart of the manufactured pieces shows a negative value in a given 10-minute period. This represents a negative value for the number of products, which is practically impossible. The explanation for this phenomenon is that a part of the manufactured batches is reworked (for example re-shaved) and these reworks do not mean new pieces of number in the production, but also the multiple machining of the same products in a given machine.

We can see that the production output was variable at the machine next to the energy consumption and because of it we refer to the next figure that shows a key indicator, the "Comparison 1. (number of pieces/kWh)". This indicator shows after the consumption of given kWh, how many product pieces can be produced by the machine (Figure 6).

We were curious during the searching of relationships and investigations that we can predict the breakdown of a given machine. We defined the process of the breakdown that an examined machine takes up more energy to the transmission of the same power. During the monitoring we found out that the OTE2 compressor consumed more kWh electricity energy during the operation of our framework as time goes by. We performed the trend calculation with exponential regression. The number of monitoring in the regression statistic was 26 666 [13]. We did not find any shortage during the fitting shortage test, that's why we can continue our work with the calculation of the determinant coefficients. From our calculation we got as a result R^2=0,84731, which points a high-level connection, because the value of R^2 had a negative influence by the increase of the sample size according to the experiences. We highlight, that the ~0,85 value is more than perfect in case of such a big data set. After that, the "best" equation of the regression parabola is the following:

$$\hat{y} = 346{,}95 * 0{,}006262^{x}.$$

As our example illustrate in practice on the horizontal axis there is a daily rating, because of it the energy consumption of the machine increases by 0,62% a day (Figure 7).

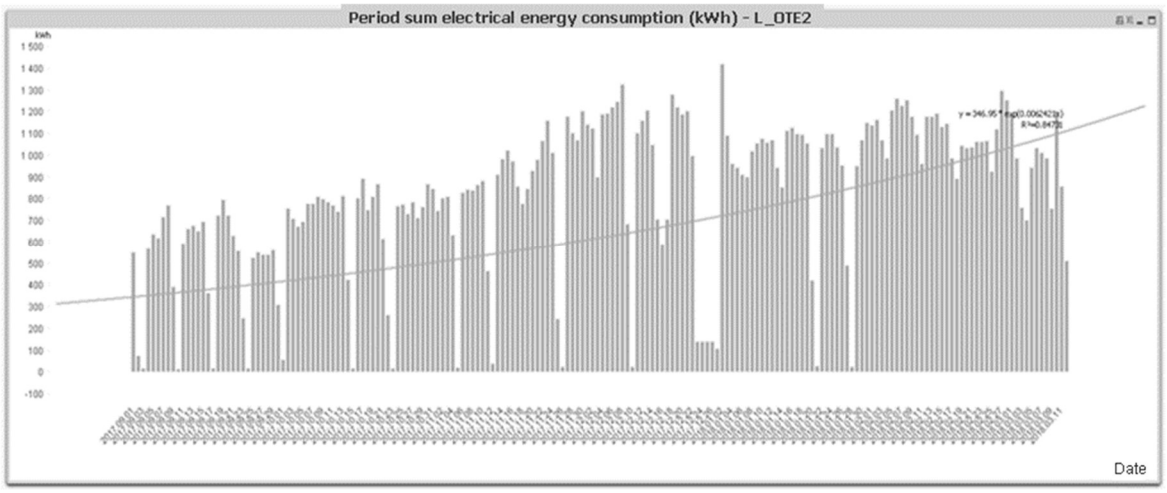

Figure 7. OTE2 equipment's daily energy consumption data and its trendline from September to March (2017-18) (Source: company's BI system)

Resource management at the company

In the previous sections we studied and analyzed the electricity management of the company, and all of energy consumption, operating and manufacturing data of each machines. Beside the energy consumption the furniture company would like to know the water and generated air consumption of the whole firm and each manufacturing machine, and the amount of air generated by compressors. To measure these data, we had to extend our framework. After a development, our system was able to measure resources' consumption data, and we could start to analyze them and make reports [19].

Beside the data measuring, our framework helps to manage the water and generated air consumption to reduce the unnecessary water and generated air usage. The optimal utilization of water and generated air is essential in the industry, and with the help of their optimization the factory can reduced their costs. The optimal consumption of water and generated air plays an important role in terms of environmental protection, because nowadays every factory puts a great emphasis on the protection of their environment and the Earth too.

In this section we introduce the measured data of water and generated air consumption of the company, and we demonstrate our results in connection with these two resources.

Monitoring of water consumption

We can say in general, that the water is the largest quantity used material in the furniture industry. The main role of water usage appears in production processes, but it can be used as helper material and it is presented in human water usage too. The water that comes into the factory we call social water. The social water is used as sewage and water vapor. A part of the water is used by the fire service for example when they have practice around the factory. We made an annual analysis the water consumption of fire service. In a given time period we detected, that the water consumption is continuously increasing, even though in the highlighted interval there was no fire service practice. We thought there were two reasons for it: the first was the broken pipe, the second was the defective sensor. The second reason was the right solution, and we changed this sensor in the system. The initial value of the new sensor was lower than the value of the old sensor, and we corrected this negative value with a simple manual overwrite. Later the water consumption of the fire service was consistent and correct [20].

One of our future goal is to optimize the water vapor consumption in the factory (see Figure 8). The waste water can be engendered in the firms by the production and in the areas, which were used by people, for example kitchens, toilets or rest rooms. As we can see, in the highlighted days (light grey mark over the columns) there was a pipe break. Because of the broken pipe the amount of used water was significantly increased.

To summarize the water consumption of the company, we can say, that the optimization is indispensable. With the help of our measuring system we can analyze the water consumption data, and we can define the optimal water consumption values in each factory area.

Generated air consumption

During this phase, we demonstrate our results in connection with the generated air consumption by compressors. We would like to know the correlation of the generated air and the electricity consumption data of a compressor. At the beginning, we don't install the sensors to the all machines, but we install one of them to the compressor housing. After that, we will install sensors to every compressor machine (Figure 9). This way we will know that one compressor how many cubic meter (m^3) of compressed air is produced by 1 kWh electricity consumption. As the Figure 10 shows, our sensors have already measured the compressed air data for all compressors

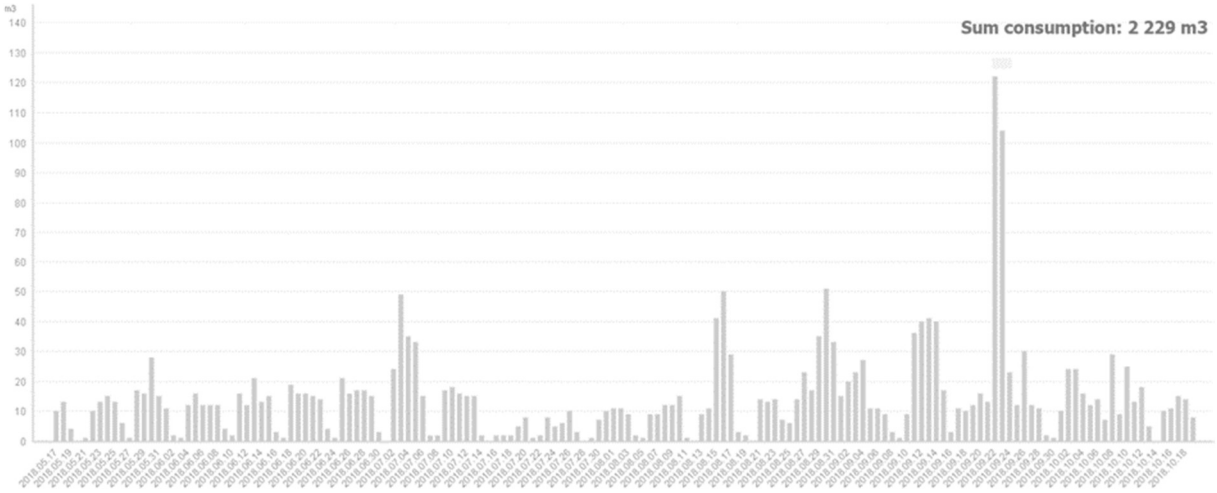

Figure 8. Waste water consumption (Source: company's BI system)

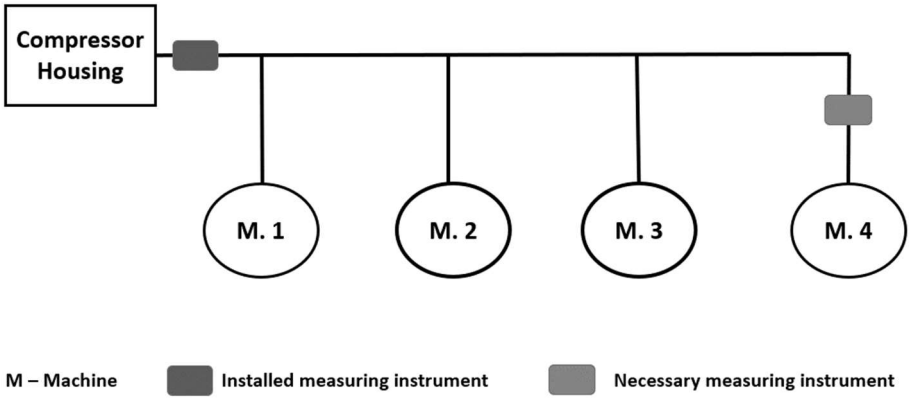

Figure 9. Compressor housing, compressor machines and measuring instrument (Source: own diagram)

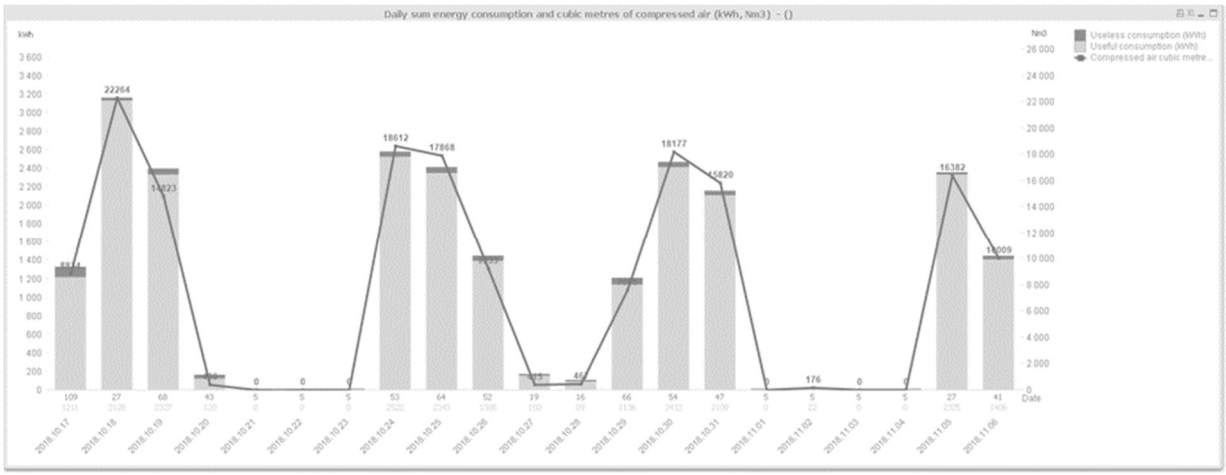

Figure 10. Electricity consumption of all compressors (Source: company's BI system)

Besides, Figure 10 shows us the performance parameters of all compressors. The dates of the time period (weeks) can be seen on the horizontal axis. The kWh energy consumption is visible on the left vertical axis. The light grey columns mark the useful, dark grey columns mark the useless energy consumptions. The numbers of generated air (m^3) are marked the right vertical axis (it's represented by the line series). Figure 12 shows an important indicator, which represents generated air m^3 / electricity consumption kWh by a compressor. We can manifest with this indicator, how effectively the compressors use electricity. In the future we will analyse the evolution of this indicator with trends, and we define a minimum value, which must fulfil the compressor. If the compressor cannot reach this minimal value intervention is needed. We can examine with the help of our filters production and consumption data of one compressor. We can go below more abstraction levels, we can examine not only the daily data, but we can measure and analyse 10 minutes data of a given compressor machine too.

Our other future goals in this field to place and install more sensors to all production machines (Figure 11). It is very important to the company to get a true and fair view of each consumption and production data. With its help they can increase not only the productivity, but they can calculate more efficiency with the monthly and annual budget too. By the numerous data measurement and data analysis the company can decrease their electricity consumption costs, and in the future, they can spend the saved money for more new development and investments.

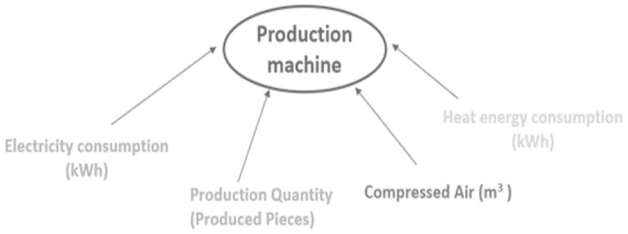

Figure 11. Measured (electricity - kWh, pcs) and still not measured (m^3, heat - kWh [21]) parameters of a production machine (Source: own diagram)

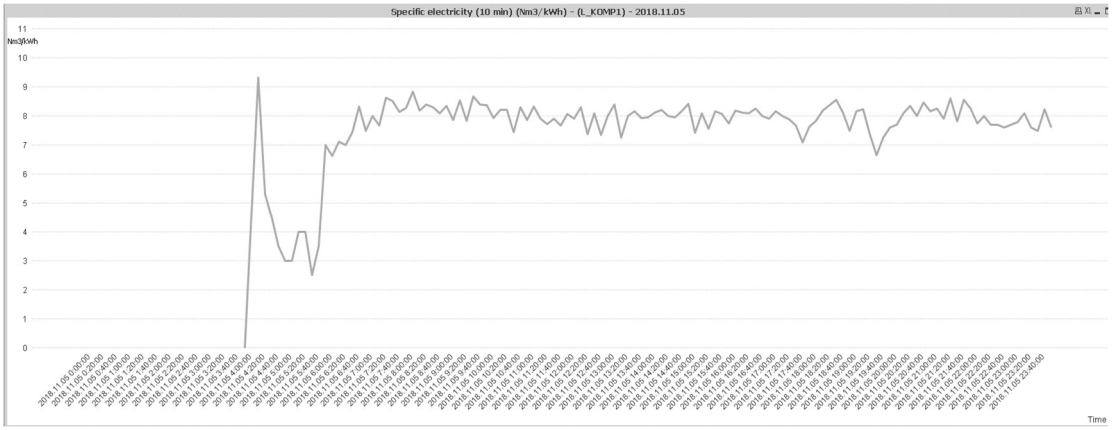

Figure 12. Specific values (generated air m^3 / electricity consumption kWh) (Source: company's BI system)

Conclusion

We configured a self-developed and integrated framework at a furniture manufacturing company. We could implement the production control and management with its help. For this, first we got to know the physical and software toolkit of the firm because the company would like to use only its own software tools to reach more efficient operation. Because of it we implemented an extended system that the company used. With this system, the data collection, correction and the data analysis were becoming possible. We made different analysis, reports and statements with the help of business intelligence system beside the live system when it has already delivered the data into our database. All of

them contained important information for the company's management and they could make more effective decisions in connection of production management. We had to prepare the system for the changes which can appear during the task in the factory, for example, when the version of the supervisory system was updated (including its internal data tables and columns too). The production controllers can expand the system too if it is necessary, the expansion of the system will no need IT knowledge.

When we examined the connection between the production and energy consumption, we identified the operating errors and the not-well used machines. We explored the problems for each machine, for example UVSOR 3, and we indicated to the managers, that this machine needs maintenance, or they need to change this machine to prevent the production stoppage. We also examined the efficiency of the manufactured products in the given machines. Here we paid a great attention to the power testing. We tested that a given machine reaches the average power value or underperforms. We made some trend analysis in connection to the energy consumption. With these trend tests we could forecast the machine failures and the repair claims (OTE2 compressor).

After the energy management and examination of electricity consumption, we dealt with the water consumption and the compressed air creation that was generated by compressors. During examination of the water consumption we discovered sensor errors and broken pipe problems. By fixing errors the unnecessary water consumption was terminated. We observed that in the hotter summer month the whole company consume more water in the personal areas (rest rooms, kitchen) and in the firms too. During our research we investigated the relationship between the electricity consumption of the compressors and the amount of air that they produce. The achievement of the compressors is important because the compressors give the compressed air to the manufacturing machines. It is very important to know how efficiently the compressors work.

Our previous developments, researches and analyses were a great help for the company. In the past the company had no opportunity for this kind of development and such depth data analysis, which enabled it to make more efficient use of resources.

In the future we would like to make more production support analysis. The development is continuous, but the production managers can make more optimal and efficient decisions with the help of the framework we developed at the given company.

Acknowledgment

Supported by the ÚNKP-19-3-3 New National Excellence Program of the Ministry of Human Capacities.

References

[1] M. Jaradat, M. Jarrah, A. Bousselham, Y. Jararweh, M. Al-Ayyoub: *"The Internet of Energy: Smart Sensor Networks and Big Data Management for Smart Grid"*, Procedia Computer Science, Volume 56, 2015, pp. 592-597, ISSN 1877-0509.

[2] M. Schulze, H. Nehler, M. Ottosson, P. Thollander: *"Energy management in industry - a systematic review of previous findings and an integrative conceptual framework"*, Journal of Cleaner Production 112, 2016, pp. 3692-3708.

[3] T. Roitsch, L. C.-Bosquet, A. Fournier, K. Ghamkhar, J. Jiménez-Berni, F. Pinto, E. S. Ober: *"Review: New sensors and data-driven approaches - A path to next generation phenomics"*, Plant Science, Volume 282, 2019, pp. 2-10, ISSN 0168-9452.

[4] Hugh Boyes, Bil Hallaq, Joe Cunningham, Tim Watson: *"The industrial internet of things (IIoT): An analysis framework"* Computers in Industry 101 (2018) pp. 1-12.

[5] M. Hermann, T. Pentek, B. Otto, Design Principles for Industrie 4.0 Scenarios: A Literature Review, Technische Universität Dortmund, 2015, pp. 11 Working paper (Accessed 12 September 2017).

[6] R. Rohankar, C.P. Katti, S. Kumar: *"Comparison of Energy Efficient Data Collection Techniques in Wireless Sensor Network"*, Procedia Computer Science, Volume 57, 2015, pp. 146-151, ISSN 1877-0509.

[7] J. Ronkainen, A. Iivari: *"Designing a Data Management Pipeline for Pervasive Sensor Communication Systems"*, Procedia Computer Science, Volume 56, 2015, pp. 183-188, ISSN 1877-0509.

[8] A. Gludovátz, L. Bacsárdi: *"Industry 4.0 projects' background and our experiences at the wood industrial manufactories"*, SEFBIS JOURNAL XI, 2017, pp. 34-41, HU ISSN 1788-2265.
[9] Z. Pödör, A. Gludovátz, L. Bacsárdi, I. Erdei, F. N. Janky: *"Industrial IoT techniques and solutions in wood industrial manufactures"*, Infocommunications Journal, Volume IX, Issue 4, December 2017, pp. 24-30.
[10] M. Persson, A. Håkansson: *"Simultaneous Data Management in Sensor-Based Systems using Metadata"*, Disaggregation and Processing, Procedia Computer Science, Volume 112, 2017, pp. 2117-2126, ISSN 1877-0509.
[11] R. Schmitt, J.L. Bittencourt, R. Bonefeld: *"Modeling machine tools for self- optimization of energy consumption"*, in: Proceedings of the 18th CIRP LCE Conference, Braunsch - Weig, 2011, pp. 253-257.
[12] A. Gludovátz, L. Bacsárdi: *"Production Related IT Solutions in the Operation of Factories"*, CINTI 2016, 17th IEEE International Symposium on Computational Intelligence and Informatics, Budapest, Hungary, Nov 17th-19th 2016, pp. 187-191.
[13] M. Mudelsee: *"Trend analysis of climate time series: A review of methods"*, Earth-Science Reviews, Volume 190, 2019, pp. 310-322, ISSN 0012-8252.
[14] A. Gludovátz, L. Bacsárdi: *"The connection of the production and the energy usage in a smart factory"*, SEFBIS JOURNAL XII, 2018, pp. 60-69.
[15] E. Cagno, A. Trianni: *"Exploring drivers for energy efficiency within small- and medium-sized enterprises: First evidences from Italian manufacturing enterprises"*, Applied Energy, Volume 104, 2013, pp. 276-285, ISSN 0306-2619.
[16] L. B. Christoffersen, A. Larsen, M. Togeby: *"Empirical analysis of energy management in Danish industry"*, Journal of Cleaner Production 14, 2006, pp. 516-526.
[17] P. Thollander, M. Ottosson: *"Energy management practices in Swedish energy-intensive industries"*, Journal of Cleaner Production 18, 2010, pp. 1125-1133.
[18] M. Schulze, H. Nehler, M. Ottosson, P. Thollander: *"Energy management in industry - a systematic review of previous findings and an integrative conceptual framework"*, Journal of Cleaner Production 112, 2016, pp. 3692-3708.
[19] Yi Man, Yulin Han, Yifan Wang, Jigeng Li, Ling Chen, Yu Qian, Mengna Hong: *"Woods to goods: Water consumption analysis for papermaking industry in China"*, Journal of Cleaner Production, Volume 195, 10 September 2018, pp. 1377-1388.
[20] Pham, Thanh Tuan, Mai, Thanh Dung, Pham, Tien Duc, Hoang Minh, Trang, Nguyen Manh Khai, Pham, Thi Thuy: *"Industrial water mass balance as a tool for water management in industrial parks"*, Water Resources and Industry, Volume 13, March 2016, pp. 14-21.
[21] J. Malinauskaite, H. Jouhara, L. Ahmad, M. Milani, L. Montorsi, M. Venturelli: *"Energy efficiency in industry: EU and national policies in Italy and the UK"*, Energy, Volume 172, 2019, pp. 255-269, ISSN 0360-5442.

Gamification in Education: Designing and Implementing a Gamified Educational On-line Tool

[1]BALÁZS BARNA, [2]TAN OMIYA, [3]YOUKI KADOBAYASHI
[1] Corvinus University of Budapest, Department of Computer Science, Hungary
[2,3] Nara Institute of Science and Technology, Laboratory for Cyber Resilience, Japan
eMail: contact@balazsbarna.hu, omiya.tan.ol1@is.naist.jp, youki-k@is.naist.jp

ABSTRACT

Gamification is the application of game mechanics and player incentives to non-game environments. When designed correctly, gamification is found to increase engagement and encourage targeted behaviours among users [3]. The paper presents the gamification of a newly created online application which goal is to help users to enhance their knowledge in a certain field. The application is created for educational purpose; therefore, the paper covers the gamification aspects related to education. In addition to this, the structure and flow of the application are also been presented. Its usability was tested with 16 university students, and the results of the test were examined in this paper.

Introduction

One of the most common reasons for poor academic performance amongst university students is poor time management due to distractions (such as work, different kind of entertainment etc.). The gamification (which is – according to Sebastian Deterding – the use of game design elements in non-game context [1]) combines a certain progress (e.g. work, study) with reward-based design aspects of games to create a product that is enjoyable and motivating as well as productive and efficient [2].

An effective game (or game-like application) has to be motivating and addictive to the users so they have a desire to use it. In addition, it has to encourage users to advance by reaching short-term goals. At the same time, a game has to maintain an option for making mistakes or failing, along with 'retrying' to the point of users succeeding in reaching their goal. [3]

The aim of the study of authors is to create an educational tool that covers a certain topic (defined by an expert or a teacher) and helps users to gain and maintain the motivation to learn the prepared materials. The desired program should work and evaluate the progress of a user anytime without continuous supervision by an expert or a teacher. The goals of this paper are to outline how an aforementioned educational tool should build up, to describe the elements that have been used for its operation, and to examine the first use of the created tool.

This paper starts with an overview of gamification itself, specifically in the aspect of education. Thereafter, it provides an introduction to the aforementioned tool, including the current state of the application, and giving early results of the first usage of this tool including the examination of its usability and effectiveness.

Throughout this paper, the terms of 'user', 'player' and 'student' refer to a person who has the willingness to learn and to use (or is made to use) a dedicated application for his/her learning process. In the following chapter, some existing gamified applications are presented in order to have a glance on how these applications can contribute to education.

Educational gamification examples

Nowadays, there is an increasing number of researches that analyse gamification. Among these researches, there are papers that examine gamified applications created for users to get familiar with a certain field [3] [4] or teach users how to work with a specific software [5] [6].

The researchers from the University of Cape Town gamified their game-development course. It aims of – among others – to improve the student's review of course material, increasing meaningful class participation, and fostering problem-solving skills. The course had a fictional story, and the syllabus followed this narrative through the semester. The story could be accessed through a website with high-quality graphic design. The students could transfer the gained points for class participation into specific options which had real-life effects (e.g. to buy an extension for their assignments). [3]

The study conducted by Domínguez et al examined the effects and outcomes of a gamified e-learning system, and they found out that gamification can have a great emotional and social impacts on students. It is because the reward systems and competitive social mechanisms might have a motivational effect on them. However, the researchers realized that if a user has difficulties to understand the operation of such application for the first time, then, it might discourage them to use the application. Moreover, if the application does not provide a completely fluid operation, and the flow is frequently stopped or interrupted by technical (e.g. errors or long process time) or administrative issues, the application will not give motivation to students to study. [4]

Several gamified applications are available for teachers (especially designed for elementary teachers but can also be used at other level of education) to enhance class participation, engagement, and motivation for learning. Such gamified applications include: Socrative, Kahoot, Classcraft, etc. [7], however, most of them require continuous supervision by an expert or a teacher, because they can give only a few-minute long tasks (Kahoot!) or they are operating as an enhanced administrative tool (Classcraft).

The next chapter provides a collection of game elements and mechanics that can be useful at a non-supervised education tool and can help to engage and motivate users to work with it.

Game elements and mechanics

There is a wide range of elements that can be applied to a gamified application. These elements can be distinguished by their aim, usability, and their effectiveness related to different behaviour patterns. About the elements, several lists were compiled [1] [8] [9], as for now, the most extensive list was created by 'Gamified UK'. The experts of 'Gamified UK' identified 52 elements and mechanics that can be used in a gamified solution [8]. (Figure 1.; [4])

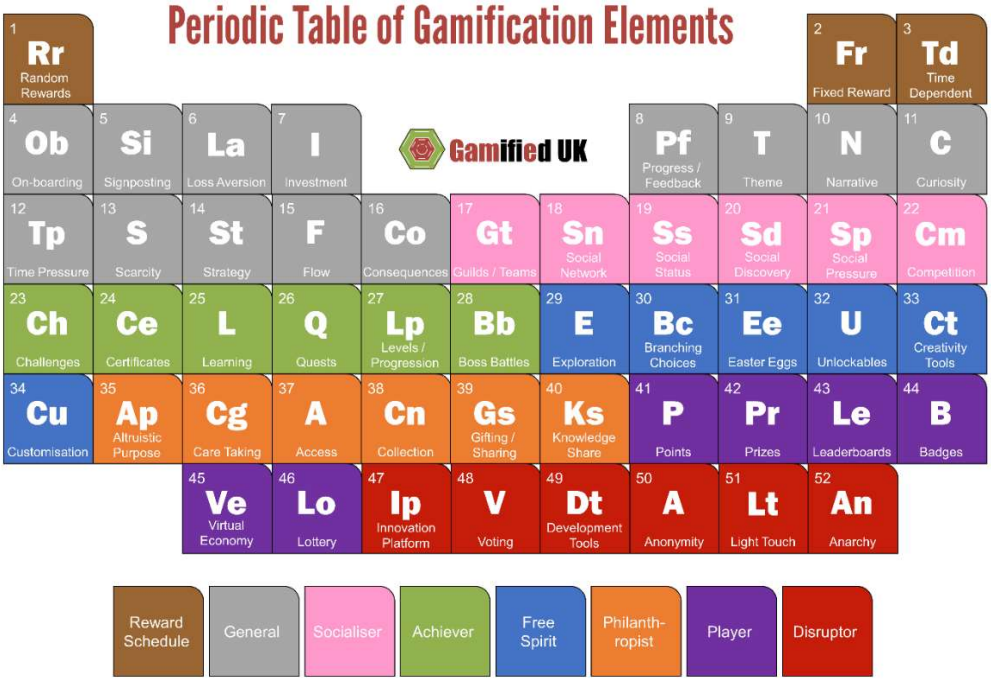

Figure 1. Periodic Table of Gamification Elements [8]

Some of the common elements are not always useful in education, e.g. as a study by Domínguez et al highlights, the leader board and process of pushing the competition between students can discourage them to participate in a gamified environment [4]. The aim of an educational tool is to improve knowledge of students, to be better than himself/herself, and not to be better than the other users. The following list (compiled by the authors) contains some of the elements that can be useful especially in the field of education:

- **Points**: the most basic element of games. For each activity such as, achieved goals, completed tasks etc. a user can gain different degree of points, which represent the amount of time and/or effort the user devoted to the service, and the points show how successful, how "experienced" the user is. [9]
- **Goals**: various objectives can be established for a user (e.g.: elaborating a topic, watching 5 videos etc.), it can affect his/her as a motivating factor. These goals can be predetermined by the system, but it is advised to maintain the possibility for a user to set the goals or select them from a list by himself/herself. This way the freedom of the user is ensured. [9]
- **Rewards**: rewards are used to honour the accomplished task, an improvement, or an activity. The rewards can be non-exclusively separated into 6 types based on their context: fixed action, random, sudden, rolling, social treasure and prize pacing. [10]
- **Consequences**: the decisions made by a user can have different outcomes, leading to a reward, a punishment, or choosing a different path of a story.
- **Feedback**: immediate response to a user's activity (or to the lack of activity) can help them to understand the context and to revise their decisions.
- **Forgive / Retry**: a gamified application which aim is to teach a certain knowledge must have the ability to 'forgive' a bad decision or an act and give a possibility to redo it.
- **Progress indicators**: such indicators can show the states of a user's goal-achieving processes (e.g.: how many points the user needs to collect to the next level), his or her activity (e.g.: how often the user visited the website in the last 20 days), the fullness of the desired skill (how many topics of the skill are completed), and then show the overall state of the 'gameplay'. [9]
- **Knowledge map**: an overview map that contains the topics that can be accomplished, and their interdependence. Every knowledge element can be a point (circle), and they can be decomposed into other sub-knowledge elements. Each item can belong to more than one sub-item, and these sub-items can contain other sub-items. Therefore, a user can decide by himself/herself which items, which routes in which order he or she would like to take. [9]
- **Story / Narrative**: a higher level of gamifying, where the knowledge-transferring and rewarding systems are restricted by a fictitious or real story. Therefore, a coherence between each task and option is stronger. For this reason, the application shows a quite consistent presence to the users, therefore they can relate to it better. It is advisable to take into consideration the field of application, the scope, and/or the environment of the potential users. [9][4]
- **Time pressure**: reducing the amount of time in which people have to complete exercises can make them more focused on a task [8]. It can also lead to different decisions, because the user does not have enough time to think through every possible outcome or to look up for the answer from external resources (e.g. internet).

The following chapter presents the flow of creating the educational tool of the authors and describes how these aforementioned game elements and mechanics appear in the created application.

Project

In a 2018 study, professor Szabina Fodor and Balázs Barna [11], examined an existing open-source e-learning service 'Moodle' to evaluate how well it can be used to gamify a course. The research showed positive results (e.g. the willingness of continuous learning week by week enhanced).

However, the boundaries of using Moodle are also visible, e.g. the service and the offered tools can be used mainly to measure the current knowledge level of students instead of helping them to learn more efficiently. Moreover, not enough improvement in its features can be applied to 'Moodle' in order to enhance the service to a highly gamified application. Due to this result, the authors decided to create an application on their own, which contains such elements, features, and flow that are described in the following chapters.

Idea

The idea of authors is to develop a web-application that teaches the users a certain field in a powerful gamified form is based on a 2014 pilot project made for a multinational company [9].

The desired tool is an online round-based game that covers the topics related to web-security. In the game, it is required (for a teacher or service provider) to implement the knowledge materials and tasks, sorting them in a hierarchical order and separate them into different topics. These materials can be activated during the gameplay by several actions such as new round, opening a map element or answering a task. The tool has been named as 'WebSec'[7].

Flow and game elements

In 'WebSec', the player starts as a newly hired employee in a company, and the player should maintain and improve his/her knowledge on web-security in order to become a 'safe' employee for the company and not to risk security attacks against the company. The aim of the game is to unlock all knowledge materials and complete all tasks with as high points as possible. [narrative, goal] The players should pay attention to several key process indicators (KPI) through the gameplay, such as the amount of *'money'* (points), the current level of company's *'security satisfaction*' towards the player and quantity of *'tolerance tokens'*. The most critical KPI is the *'security satisfaction'* which is a percentage number. If it reaches zero, the company loses all confidence towards the player, and the game is over. The higher the value, the more reliable the player is in the company. The *'tolerance tokens'* help to reduce the negative effect of a bad decision. As the tokens run out, the level of satisfaction is more vulnerable which means that the company tolerates less a bad decision and the *'security satisfaction'* level reduces in a higher volume. The aim of *'tolerance tokens'* is to forgive occasional mistakes but punish continuous bad decisions (as a sign of lack of required knowledge). (see Figure 1.) [rules, points, KPI, forgive/retry]

After starting the game, the user finds himself on a company map with clickable elements, can see the current state of KPIs and a time field where the flow of days can be checked. The map elements can open topic specific libraries or can lead to another map. [exploration, knowledge map]

Figure 2. A wrongly answered task with displaying the correct and wrong selections, and with the received punishment in the top right corner (own designed)

At the beginning of every round, the *'security satisfaction'* and *'money'* reduces a bit, maintaining the urge to learn more and solve more tasks. Based on the 'activated' knowledge materials the game can give a user tasks or new materials will become accessible. In the current state, the available types of the tasks are single/multiple choice and text-fill questions, extended with short questionnaires that work like case-studies. [time, flow, learning, task].

As for the tasks, the player always receives feedback whether a given answer is correct or not. Depending on the correctness, a player can receive rewards

[7] http://websec.infora.hu

(increased KPIs) or punishment (decreased KPIs). If the player could not answer correctly, then in the next rounds the same task appears again until the player gives a correct answer (see Figure 2.). A correctly answered task may lead to unlocking new knowledge materials. [feedback, consequences, rewards].

Unlocked knowledge materials can be accessed in the knowledge library that contains the activated articles separated by topics. The number of available and all materials on a specific topic are displayed in order to check how many more articles are needed to unlock. [knowledge, progression]

Technical details

The application was created in PHP language at the back-end side with using the framework of Symfony3. The front-end has been created with using HTML, CSS, and JavaScript, using the library of jQuery. As for handling its data MySQL database is used. The background images are designed by iconicbestiary / Freepik via freepik.com.

Expectations for the first version

In order to determine the proper functioning of the first version of the application, we have identified some expectations. The first two expectations examine the usability of the application, while the third expectation improves the efficiency of application enhanced by gamification.

- Firstly, a user has the willingness to play a gameplay for at least 60 minutes.
- Secondly, after restarting the application, a user can play for a longer time than the previous gameplay.
- Thirdly, a user can learn a curriculum better with the application in one-unit time than in a lecture led by an expert using presentation.

The application has been tested once by Japanese students. During this test, the usability and the usefulness of the application, and whether the application has met the aforementioned expectations are examined.

Practical usage

The application has been tested in two groups, each contains 8 university students from Japanese universities[8]. The students had limited time to use the application, and they used it at the same time period.

The 16 users started 43 gameplays, 26 of them had a duration of more than 5 rounds. It is impossible to complete the game within 5 rounds, neither with success nor with failure, therefore those gameplays with less than 5-round duration are counted as non-serious try. Hereinafter, only those gameplays are counted that are longer than 5 rounds.

Duration

Every student had at least one gameplay with more than 5-round duration, on average, they played 1.63 times. The average length of these sessions is 72.65 rounds with a standard deviation of 45.63 rounds.

The time invested in the gameplays varies. The shortest session was 3-minute long, the longest lasted for 95 minutes, 16 of them (62%) lasted more than 30 minutes, and 12 gameplays (46%) reached the 1-hour duration (Figure 3, A). The average time spent is 47 minutes with the deviation of 29 minutes.

It can be seen that the application can be used for a longer period, such as 30-60 minutes, however the occasion was a situation supervised by teachers, and the main task for the students was to use this application. As for 'retry', 6 users started new gameplays overall 10 times. 8 gameplays out of 10 had a longer duration, and they were averagely 20 minutes longer than their previous gameplay (Figure 3, B).

Progression

In the application, 198 knowledge articles have been implemented, separated into 10 topics. In 77% of gameplays, less than half of the articles have been activated, however one player could unlock 98% of knowledge materials.

[8] Nara Institute of Science and Technology (Ikoma) – 10, Institute of Information Security (Yokohama) – 3, Keio University (Tokyo) – 1, Osaka University (Osaka) – 1, Kyoto University (Kyoto) – 1

Throughout the gameplays, 124 different tasks have been activated. The players answered these tasks overall 1266 times, and 67% of the cases were correct which means the players could give the correct answers in their first try in more than half of the cases.

Approximately, the third of tasks has been answered incorrectly. The aim of the tool is not to use those kinds of tasks that stop the users in the flow. However, if the tasks are too simple, they do not measure the particular knowledge material correctly.

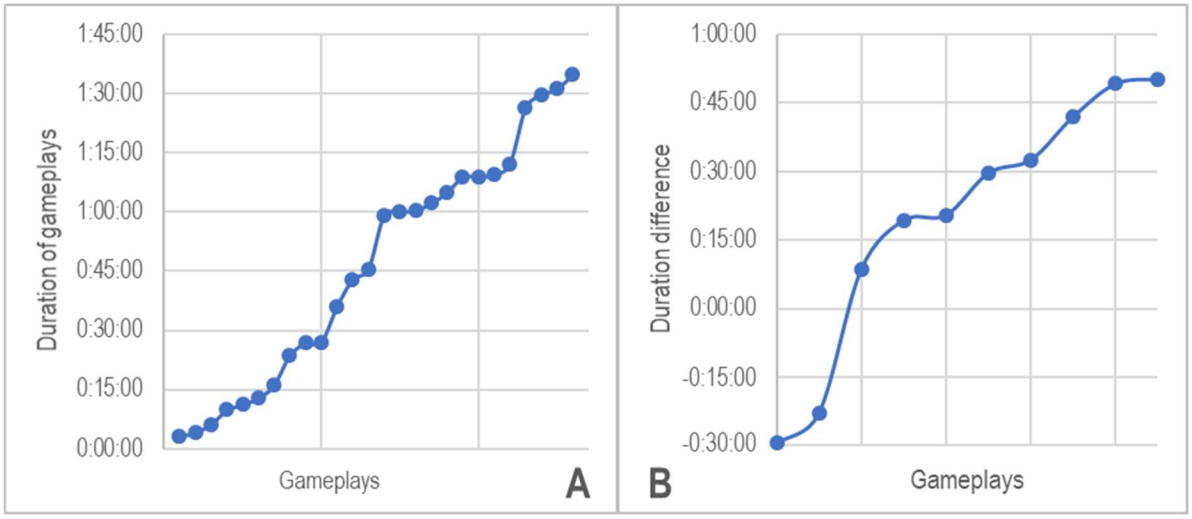

Figure 3 **A)** Gameplays ordered by their time duration; **B)** Restarted gameplays ordered by difference between time duration of 2 consecutive gameplays played by the same user (Own figure, 2019)

Evaluation of usage

We defined three expectations. The first one refers to its time-related usability, specifically, whether the application is usable for an at least 60-minute long gameplay. The results show that 46% of gameplay lasted more than 60 minutes, however, the occasion was a supervised situation, thus the users probably had a slight external pressure to use the application. Therefore, it is possible to use the application for at least 60 minutes, but it only proves the technical usability of it, not the engaged usage of a user. The expectation needs further verification in a less supervised environment.

The second expectation refers to retrying, whether the following gameplay has a longer duration than the previous one. In those cases where a user started a new gameplay after failing one, 80% of them had a longer duration with an average 20-minute surplus. Due to these facts, it can be stated that the application has met the second expectation.

In order to verify the first expectation, a longer time period is needed. At the next test, the users attend an introductory session within a lecture. Following the lecture, the users have 1 week to use the application, after that time period, the proper measurements about the engagement of users can be taken. There can be a second session where the top achievers are highlighted and/or rewarded, increasing their motivation to use the application.

Further research is required to examine the third expectation where a control group is granted. The control group would take a lecture in the field of web-security, and another group would use the application for the same time period.

Conclusion

This paper elaborates the current phase of the research of authors which is to create an online gamified tool that helps a user to master a particular topic. After a brief introduction, a few game elements that can be useful in the field of education were described.

The idea and structure of 'WebSec', the web-security training gamification tool were presented. After describing its gamification related features, the result of the first usage was examined. The outcome shows that a user can play with 'WebSec' for more than 60 minutes in a supervised environment, however, the engagement factor of the application needs further verification in a less supervised situation. As for re-playability, with almost every following try, the player can complete more tasks and unlock more knowledge materials, therefore it can be counted as a sign of improving in the given field.

Currently, it is not yet verified by the authors whether the gamified elements of the tool help the user to learn more. Also, the willingness to use the tool without supervision needs additional verification. Further feature-, stability- and reporting-improvements are going to be implemented by the authors which aim – among others – to analyse better the progress of user during the gameplay, to detect possible idle state during of its usage, and to give the ability to personalise for users.

Although the tool requires additional improvements and further usage tests, the current results are promising, and it is advised to do additional research.

Acknowledgement

This article has been supported by the ÚNKP-18-3 New National Excellence Program of the Ministry of Human Capacities.

References

[1] Deterding, Sebastian – Dixon, Dan – Khaled, Rilla – Nacke, Lennart (2011): From Game Design Elements to Gamefulness: Defining "Gamification", MindTrek '11 Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, pp. 9-15.

[2] O'Donovan, Siobhan (2012): Gamification of the Games Course, at: http://pubs.cs.uct.ac.za/archive/00000771/01/Gamification_of_the_Games_Course, date: 2019.01.11.

O'Donovan, Siobhan – Gain, James – Marais, Patric (2013): A case study in the gamification of a university-level games development course, SA-ICSIT '13 Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference, pp. 242-251

[3] Domínguez, Adrián –Saenz-de-Navarrete, Joseba – de-Marcos, Luis –Fernández-Sanz, Luis –Pagés, Carmen –Martínez-Herráiz, José-Javier (2013): Gamifying learning experiences: Practical implications and outcomes, Computers & Education, vol 63, pp. 380–392.

[4] Li, Wei – Grossman, Tovi – Fitzmaurice, George (2012): GamiCAD: A Gamified Tutorial System For First Time AutoCAD Users, In Proceedings of the 25th annual ACM symposium on User interface software and technology, October 7-10, Cambridge, Massachusetts, USA, ACM, pp. 103-112.

[5] Dong, Tao –Dontcheva, Mira –Joseph, Diana – Karahalios, Karrie – Newman, Mark W. –Ackerman, Mark S. (2012): Discovery-based Games for Learning Software. In Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems, May 5-10, Austin, Texas, USA, ACM, pp. 2083-2086

[6] The Tech Edvocate (2017): 8 must have gamification apps, tools, and resources, at: https://www.thetechedvocate.org/8-must-gamification-apps-tools-resources/, date: 2019.01.13.

[7] Gamified UK (2017): 52 Gamification Mechanics and Elements, at: https://www.gamified.uk/user-types/gamification-mechanics-elements/, date: 2019.01.10.

[8] Barna, Balázs (2016): Gamification in Education. In SEFBIS Journal Vol. 10, pp. 66-74., John von Neumann Computer Society – SEFBIS, Budapest

[9] Yu-kai Chou (2013): The Six Contextual Types of Rewards in Gamification, at: https://yukaichou.com/marketing-gamification/six-context-types-rewards-gamification/, date: 2019.01.11.

[10] Barna, Balázs – Fodor, Szabina (2018): An Empirical Study on the Use of Gamification on IT Courses at Higher Education. In: Auer M., Guralnick D., Simonics I. (eds) Teaching and Learning in a Digital World. ICL 2017. Advances in Intelligent Systems and Computing, vol 715. Springer

SEFBIS Board's Adopted and Performed Decisions of 2018

Decision No. 1/2018.

The GIKOF SIG Board discussed and accepted the Annual Workplan required by the Secretariat of the. JvN CS Decisions made on 11th of January, 2018:

- Acceptance of the Annual Report about activities in 2017 by the JvN CS Leadership
- Review and validate the Board and the SEFBIS SIG membership
- Actualize the SEFBIS Website: http://raffa6.wixsite.com/sefbis
- Cooperation in organizing the 15th OGIK Conference in Sopron (November 2018)
- Upload the SEFBIS Journal to the database of EBSCO for indexing
- Cooperation with other SIG groups of JvN CS
- Decision on the 16th OGIK/ISBIS Conf. 2019

Decision No. 2/2018

SEFBIS Board meeting at the General Assembly of JvN CS and made decisions as follows:

- Conference *date* and *venue*: 9-10. November 2018 Sopron University
- The International Programme Committee:
  Chair: RAFFAI, Mária (Széchenyi István University); *Vice-chairs:* DOBAY, Péter (Pécs University of Sciences) and BACSÁRDI, LÁSZLÓ (Sopron University)
  Invited Members: CHROUST, Gerhard (Johannes Kepler University, Austria); DOUČEK, Petr (Prague University of Economics, Czeh Republic), GÁBOR, András (Budapest Corvinus University); KŐ, Andrea (Budapest Corvinus University); KRUZSLICZ, Ferenc (Pécs University of Sciences); TJOA, A Min (Vienna University of Technology, Austria) and UCHIKI, Tetsuya (Saitama University, Japan)
- The *Chair of Organizing Committee* of the OGIK'2018: BACSÁRDI, LÁSZLÓ (Sopron University)
- *Sponsors* of the Conference will be: JvN CS, Sopron University, Foundation Alexander, Guidance Ltd.

Decision No. 3/2018

SEFBIS SIG Board meeting held on 10th of November 2018 after the 15th ISBIS Conference

- The SEFBIS SIG Board expresses gratitude and thanks to organizers, reviewing board members and authors to successful work of the annual 15th OGIK/ISBIS Conference.
- Decision about the selection of the papers that will be published in the SEFBIS Journal No. 13 (in English) and about the date of sending the papers to the Committee by 28th February 2019 the latest. The papers, the contact with the authors, the reviewing process; the editing, printing work will be managed by Maria Raffai.

Decision No. 4/2018

The SEFBIS Board revised the members' activity and renewed the membership. The annual Activity Report on 2018 and the Workplan of SEFBIS SIG for 2019 have been prepared. Decision was made about the circulation in December and sending to the Secretariat of the JvN CS by 20th December, the latest.

Report on ISBIS/OGIK 2018

9-10 of November, Sopron

International Forum of Scientific and Educational Forum on Business Information Systems

The SEFBIS/GIKOF Board revised the applications for organizing the 15th conference and decided to give the opportunity again to the Sopron University. The colleagues at the university were not only very much committed in planning, managing and performing the OGIK conference in Sopron, but they had already experiences from the last year. They found, that the new solutions, new forms of presentations the they introduced gained the participants over the usefulness, so they could make the event more effective and attractive. László Bacsárdi, the chair of Organizing Committee and his eager colleagues such as Gergely Bencsik, Attila Gludovátz, Péter Kiss, László Koloszár, Zoltán Pődör and Mónika Tóth have been constantly ready to arrange everything in connection with the program. Nevertheless, not only the Organizing Committee worked hard during 8 months before the conference, but also the members of the International Program Committee, who invited reviewers and also reviewed the papers, and managed the professional program. Anyway, the hard work finally fructified, the authors have sent interesting and valuable papers, and the presentations arouse a high interest on the different topics.

The 15th OGIK Conference was organized on 9-10th of November 2018 in Ligneum Conference Centre of the Sopron University. The conference Call for Papers inspired the authors to submit works dealing with results and achievements in the field of ICT innovation. Members of the Program Committee have controlled altogether 68 subscriptions with a double-blind review process. They have accepted 28 papers for conference presentation and 17 for poster presentation. The speeches were organized in one plenary and 4 main sessions, such as: Research and Development (in English), Information Management (in English), Interdisciplinary Session both in English and in Hungarian), Internet of Things (in Hungarian), there was a Round table discussion and a "One minute Madness" Poster presentation session with joined wine tasting.

The "*plenary session"* was introducing by salutation of the Sopron University's rector, András NÁHLIK and the chair of SEFBIS/GIKOF, Mária RAFFAI.

The plenary session was continued by speeches of acknowledged professors:

- Tetsuya UCHIKI (Japan) was talking about an information systems design approach that is suitable to the social context in Japan.
- A Min TJOA (Austria) emphasized the importance of sustainability for business information systems, while
- Hiroki TOMIZAWA and Akihiro Abe (Japan) implemented an analysis of a digital archive system implementation that is based on design of utilization model of Earthquake-related material.

The afternoon and the Saturday program offered different interesting sections to the audience:

- "*Section #1: Research and Development*" This section aimed to present and discuss topics that are related to scientific results and the application development works which support the business processes and help the users in their every day work in order to perform their tasks faster, more effective and precise or even automatically. The main topics focused on the energy management system, on optimizing energy consumption, the authors showed us solutions for solving exercise generation problems, for forecasting performance improvement, and we could even learn how can the engineers reduce the energy cost in Jordan. on technologies for creating IoT infrastructure, on agent-based simulation software for application management and on the newest way for using ERPs.
- "*Round Table Discussion: Characteristics of the profession of BIS*" Attila Fekete (FekiWebstudio Kft.) has run this discussion with inviting colleagues who are involved in this field of IT. The audience was keen on the questions and the answers, and also joined to the speakers with expressing their thoughts and mind.
- *"Section #2: Information Management"* has ever been a hot topic for our forums. We could hear interesting presentations about ICT solutions, variety of eServices, business value of decision support systems, about the knowledge management, about the users' requirements and behavior, the maturity of information security and about other interesting topics. The presentations were followed questions from the audience and discussions.
- The *"One Minute Madness" poster presentations section* focused on different topics that the authors are dealing with. All authors got 1-minute time for presenting their topic and calling the attention of the audience, who could later give questions from the authors in front of the poster-table. This kind of "poster-style" was really interesting and the participants found it very effective. The authors challenged the audience and called them for a detailed discussion about different topics, such as Big Data, interpretative methods for IS evaluation, analysis results from different fields, cloud computing serving organizational culture and leadership, tools and methods for efficient project management, Cyber threats and protection, digital transformation in EU states, analysis and impact of Industry 4.0, IoT tools etc.

- *"Section #3: Interdisciplinary"* was organized on topics that cover at least two different scientific fields and or ICT and a branch of business/economy. The speakers called our attention to the importance of impact and value of ICT, they were talking about the ICT problems in government, in insurance companies, in healthcare, but it was also emphasized how important to analyze the distances and costs of economic networks or what opportunities we have in checking community media contents.
- *"Section #4: Internet of Things"* In the most interesting presentation the author was talking about an actual problem: How can the Z-generation assert oneself on labour market in Industry 4.0 revolution? The other presentation that piqued the interest, was dealing with the artificial intelligence' impact on the society.

After the conference we conclude that the colleagues deal with the most up-to-date problems not even of ICT but also the interdisciplinary ones and do not spare pains to find the best solutions for solving the problems. The speakers and participants reported on problems that have to be solved, on new technologies, solutions that serve the innovation. Consequently, we need urgent actions to save our environment, and the conferences like the OGIK and the common thinking help us to reach our goals! Participants closed the Conference with acknowledgement for the smooth management and hospitality of Sopron University and for all the colleagues who played active role in organizing, managing and performing the OGIK'2018 conference!

Mária Raffai
Chair of the IPC

Conferences Organized Worldwide and Relevant to the SEFBIS Community

Event	Date/Location	Organizers' Contact
The IFIP 20th **Open Conference on Decision Support**	24/06-26/06 2020 Wroclaw, PL	zygmunt.mazur@pwr.edu.pl Tel. +48713202019
I3E 2020, e-Business, e-Services, and e-Society	06/04-08/04 2020 Skukuza, ZA	machdel.matthee@up.ac.za Tel. +27 12 420 3365 Fax +27 12 420 5346
International Conference on **Interoperability for ERPs and Applications**	26/03-27/03 2020 Tarbes, FR	Bernard.Archimede@enit.fr Tel. +33 5 62 44 27 34
CONFENIS International Conference on Research and Practical Issues of Enterprise Information Systems	17/01-18/01 2020 Sharjah, AE	xiaobo@aus.edu Tel. +971 50 296 0815 Fax +971 6 558 5065
The First International Conference on **Industrial Information Integration Engineering**	16/01-17/01 2020 Sharjah, AE	xiaobo@aus.edu Tel. +971 50 296 0815 Fax +971 6 5585065
Empowering teaching for **digital equity and agency**	06/01-11/01 2020 Mumbai, IN	amina.charania@tiss.edu Tel. +91-7588699650
CONFENIS 2019 – International Conference on Research and Practical Issues of Enterprise Information Systems	16-17/12/ 2019 Prague, CZ	IFIP WG 8.9 petr.doucek@vse.cz
PoEM 2019: The Practice of Enterprise Modelling [04333]	27/11-29/11 2019 Luxembourg, LU	e.proper@acm.org Tel. +352621762041
ISBIS–OGIK 2019 16th International Symposium on Business Information Systems	8-9 /11/ 2019 Budapest, HU	NJSZT GIKOF SIG raffa6.wix.com/sefbis
2nd IFIP International Cross-Domain **Internet of Things** Conference	31/10-01/11 2019 Tampa, Florida, US	strous@iae.nl Tel. +31641767045
I3E 2019 eBusiness, eServices, eSociety	18-20/09 2019 Trondheim, NO	IFIP TC6, WG 6.11 ilpappas@ntnu.no
INTERACT 2019 – International Conference on Human-Computer Interaction	02-06/09 2019 Paphos, CY	IFIP TC13 pzaphiri@gmail.com
MIM 2019 – 9th IFAC Conference on Manufacturing Modelling, Management and Control MIM	28-30/08 2019 Berlin, DE	IFIP TC5, WG 5.7 divanov@hwr-berlin.de
DisCoTe – International Federated Conference on Distributed Computing Techniques	18-21/06 2019 Kongens Lyngby,	IFIP TC6, WG6.1 albl@dtu.dk
WiOpt 2019 – The 17 International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks	27-31/05 2019 Avignon, FR	IEEE Control Systems Society elazouzi@univ-avignon.fr

SEFBIS Action-Plan for 2019

The Board of SEFBIS Special Interest Group

Chair: Mária RAFFAI
Members active in 2018: András GÁBOR, Zsolt KOSZTYÁN, Andrea KŐ, László BACSÁRDI, Ferenc TÓTH
Honorary members: Péter DOBAY, Gábor HOMONNAY

At the first meeting held in January 2018 the Board decided to perform the following tasks for the Year 2019:

Description	Estimated deadline	Estimated venue	Participants, responsibility
Discuss, plan and accept both the action- and financial-plan for 2019	Until 15 January 2019		Members of the SEFBIS Board
Active participation at the meeting of the JvN CS professional communities, giving proposals for more effective cooperation with Secretariat and the other SIGs.	February 2019	Meeting Room of JvN CS	SEFBIS SIG Chair
Accepting papers to SEFBIS Journal No. 13	28th February 2019 latest	NA	Editor in Chief
Reviewing process for paper having sent to SEFBIS Journal	31st May 2019	NA	Editor in Chief and reviewers
Managing the work of reviewing, contacting with authors and arranging editing process and publishing the professional Journal(s).	continuous	NA, virtual	Approximately 12 reviewers
Update the GIKOF/SEFBIS website (content and design)	continuous	NA	SEFBIS Chair
OGIK/ISBIS'2019 conference: call for papers, applying for sponsorship, advertising, activating the Conference Management System (EasyChair), reviewing papers, organizing work	from April to November 2019	NA, virtual	12 colleagues, SEFBIS Board, invited IPC, experts
Cooperation both with leaders of business and university departments on BIS	continuous	face to face and virtual comm..	4-5 Board member Resp. SEFBIS pres.
Managing the editorial and printing work of the SEFBIS/GIKOF Journals; Uploading the new Journals to the international EBSCO Database	continuous	NA, virtual	Editor in Chief and Chair
Development and uploading database on experts, professionals in the field of BIS in Hungary	continuous	gathering information	2-3 Board members + activists, students
Preparing a competence map of professionals/teachers, Organizing on-line lectures for all BIS students in Hungary	continuous		2-3 Board members + professors
OGIK/ISBIS'2019 Conference	8-9th November 2019	Budapest, Milton Friedman Universiity	60-80 participants + students, 40-45 papers
Evaluation of the yearly activity: results, report to JvN CS Presidency	18 December 2019 latest	NA, virtual	SEFBIS SIG Chair

International Conference on Research and Practical Issues of Enterprise Information Systems

CONFENIS 2019 provides an international forum for Enterprise Information System (EIS) researchers and practitioners from all over the world to come together, present and discuss their latest research findings and ideas. The conference is specifically aiming at facilitating the exchange of ideas and advances in all aspects and developments of EIS. CONFENIS 2019 invites EIS-experts who are interested to present and disseminate their work at an international forum to submit their contributions. The proceedings of the conference will be published by Springer LNBIP.

The Venue:
The CONFENIS 2019 will be held at **Prague University of Economics**, Czech Republic on **16-17 December 2019**

The proposed but not limited topics of CONFENIS 2019 are the followings:

- Business Modelling and Business Process Management
- Modelling Formalisms, Languages and Notations
- Model Driven Architectures and Engineering
- EIS Concepts, Cases and Management
- Governance in EIS, Business Value of EIS
- Supply Chain Management Aspects
- Decision Support Systems
- Enterprise Knowledge Engineering and Management
- Databases and Information Systems Integration,
- Big Data, Data Mining
- Ontology-based Data Integration
- Open Source EIS
- Cloud Computing Services
- Industry 4.0 and Artificial Intelligence
- Internet of Things Applications
- Security and Privacy Issues
- Human-Computer Interaction on EIS
- Collaborative and Social Interaction
- Workflow and Project Management
- EIS Teaching and Education

All papers must be original and not simultaneously submitted to another journal or conference. Full papers and Poster paper categories are welcome: Registration available through https://easychair.org/conferences/?conf=confenis2019

International Program Committee
General Chairperson: Petr DOUČEK (Prague University of Economics – Czech Republic)
Chairpersons: A Min TJOA (Vienna University of Technology – Austria), Maria RAFFAI (Szechenyi University – Hungary), Josef BASL (Prague University of Economics – Czech Republic)
Organizational Coordination Chairperson: Antonin PAVLIČEK (Prague University of Economics – Czech Republic, antonin.pavlicek@vse.cz)

For further information on the WCC'2019 please visit the conference website:
https://infosec-conferences.com/events-in-2019/confenis/

ISBIS / OGIK Conference'2019

8-9. of November 2019; Budapest

16th Conference on Business Information Systems

The SEFBIS (GIKOF) Special Interest Group of the John von Neumann Computer Society organizes its conference on Business Information Systems already 16th time. This event is followed with great interest of both the professionals and the users. from around the world. The conference is a great opportunity for the domestic and foreign experts, developers, students and users to represent and discuss their results, to change their minds and get together face to face. Emphasizing the important role of women in IT-sector the ISBIS'2019 Conference gives opportunity to the NJSZT NOKIT community to organize a whole session.

The planned topics that are not limited to:

- The business value of ICT, methods of controlling efficiency and of analysis
- New technologies supporting business innovation and competitiveness
- Business process modeling; model driven architectures, open source solutions
- System engineering concepts, methods and tools, modeling languages
- Knowledge based systems in business, efficiency; business intelligence
- The AI solutions in business and the impact on the society
- The increasing importance of IT security, effective solutions
- Big Data management: storing and processing, data mining; analysis
- Women in IT sector: increasing role, appreciation at workplaces
- Industrial solutions: challenges of Industry 4.0 and IoT – software and architectures
- Enterprise wide integration → integration of data, database and applications
- Multidisciplinarity: ICT applications medicine, self-driving cars, utilization of visualization etc.
- Education: CSc, MSc courses on business information systems, trainings, labour market needs vs teaching curriculums, changing competences, change management in education, teaching methods and tools, mobility etc.

The official language of the conference is both Hungarian and English. The authors have to present their results in the language of the abstract. The authors and speakers are asked to send an extended abstract in max 4,000 characters. The IPC offers different types of presentations:

- live presentation in 15-20 minutes and
- poster demonstration

The abstract should be sent through Easychair CMS in English or in Hungarian; on https://easychair.org/account/signin?l=eg0dBVfA1z2AV0HXt97kBT The information: on the GIKOF or the conference Web Page: https://uni-milton.hu/16-ogik-orszagos-gazdasaginformatikai-konferencia/. It is the IPC's responsibility to let the abstracts review and decide about the acceptance.

International Program Committee and Conference Organizers

Chair of the IPC: Maria RAFFAI (Széchenyi István University, Győr)
Vice Chairs: Péter DOBAY (Pécs University of Sciences)
Chair of the Organizers Committee: Vid Sebestyén Honfi (Milton Friedman University)

Further details:

in Hungarian: uni-milton.hu/16-ogik-orszagos-gazdasaginformatikai-konferencia/ or raffa6.wixsite.com/ogik
in English: raffa6.wixsite.com/sefbis/ogik-2019

Table of Content


Scientific and Educational Forum on Business Information Systems
S·E·F·B·I·S