INFORMATION SYSTEMS IN MANAGEMENT

Systemy informatyczne w zarządzaniu

Vol. 3 **2014** **No. 1**

Quarterly

Information Systems in Management
Primary version of the journal is an electronic version.

Address of the Editor
Faculty of Applied Informatics and Mathematics, WULS – SGGW
ul. Nowoursynowska 166, 02-787 Warszawa, Poland
e-mail: isim@sggw.pl, www.isim.wzim.sggw.pl

ISSN: 2084-5537

Wydawnictwo SGGW
ul. Nowoursynowska 166, 02-787 Warszawa, Poland
e-mail: wydawnictwo@sggw.pl, www.wydawnictwosggw.pl

Print: Agencja Reklamowo-Wydawnicza A. Grzegorczyk, www.grzeg.com.pl

INFORMATION SYSTEMS IN MANAGEMENT

Vol. 3 **2014** **No. 1**

Table of contents

Katarzyna Dohn, Adam Gumiński, Mirosław Matusek, Wojciech Zoleński
IT TOOLS IMPLEMENTED IN THE INFORMATION SYSTEM SUPPORTING MANAGEMENT IN THE RANGE OF KNOWLEDGE MANAGEMENT IN MECHANICAL ENGINEERING ENTERPRISES 3

Dariusz Doliwa, Wojciech Horzelski, Mariusz Jarocki
THE SYSTEM OF PLANNING AND MANAGEMENT OF THE MEDICAL SERVICES 16

Marcin Hernes, Jadwiga Sobieska-Karpińska
CONSENSUS DETERMINING ALGORITHM FOR SUPPLY CHAIN MANAGEMENT SYSTEMS 27

Piotr Milczarski, Zofia Stawska
COMPLEX COLOUR DETECTION METHODS USED IN SKIN DETECTION SYSTEMS 40

Michał Henryk Rudowski
MANAGEMENT INFORMATION SYSTEMS IN POLISH RAILWAY TRANSPORT IN THE CONTEXT OF THE TECHNOLOGICAL AND THE ORGANIZATIONAL CHANGES 53

Jacek Woźniak, Piotr Zaskórski
SECURITY MANAGEMENT IN THE *H-H TYPE* OF THE ORGANIZATIONAL STRUCTURES 64

Danuta Zakrzewska
ON USING DATA MINING TECHNIQUES FOR CONTEXT-AWARE STUDENT GROUPING IN E-LEARNING SYSTEMS 77

INFORMATION
SYSTEMS IN
MANAGEMENT

Information Systems in Management (2014) Vol. 3 (1) 3−15

IT TOOLS IMPLEMENTED IN THE INFORMATION SYSTEM SUPPORTING MANAGEMENT IN THE RANGE OF KNOWLEDGE MANAGEMENT IN MECHANICAL ENGINEERING ENTERPRISES

KATARZYNA DOHN, ADAM GUMIŃSKI, MIROSŁAW MATUSEK,
WOJCIECH ZOLEŃSKI

Faculty of Organisation and Management, Silesian University of Technology

In the paper the authors showed and characterized IT tools implemented in the information system supporting management in the range of knowledge management in mechanical engineering enterprises. The authors gave a detailed description of the implementation of the following IT tools: knowledge objects management, risk management, case based reasoning.

Keywords: information tools, knowledge management, mechanical engineering enterprises

1. Introduction

Dynamic development of economy requires more and more effective solutions for knowledge management. This relates both to the concept of knowledge management supporting models and appropriate technological infrastructure which enables implementation of those solutions. Enterprises use information systems to support business processes based on quantitative collection and processing of information. However, effective knowledge management requires selective choices of knowledge sources, appropriate collection procedures of them, their transfer and use in the course of business procedures.

In case of the mechanical engineering industry that executes contracts for individual orders of their customers [3, 4, 5], it is necessary to support knowledge management in the complex system of internal processes [1, 2]. Access to contextual, quality knowledge for authorized users is one of the key factors which decide about the success of the contract portfolio in the mechanical engineering industry.

This article presents IT tools implemented in the Content Management System in the range of Knowledge Management designed on the basis of survey research. It also presents characteristics for each of the used IT tools and the scope of use for the support of particular knowledge processes. Establishment of a management support model for knowledge management and construction of an IT system on that basis were the main goals of the research and development project called "An IT system supporting knowledge management in enterprises of the mechanical engineering industry" run in the Faculty of Management and Administration of the Silesian University of Technology.

2. Characteristics of the it tools implemented in the knowledge management supporting system in a mechanical engineering enterprise

Contracts which require customized approach to meet the demands of a customer require an individual approach which uses resources of an enterprise in such a way as to provide the best possible support of the business processes. Effective collection of knowledge and its processing within the designed system required implementation of the following tools to support the knowledge processes:

- Case Based Reasoning (CBR),
- Knowledge Object Management System (KOMS),
- Risk Management Support Tool (RMST),
- Human Resources Support Tool (HRST),
- Rule-based Expert System (RES),
- Interactive Pair-matching System (IPS).

Detailed characteristics of those implemented tools can be found in the subsequent chapters.

2.1. The concept of Case Based Reasoning (CBR) in a management supporting system in the range of knowledge management

Case Based Reasoning uses specific knowledge of experience from past situations called "cases". The CBR method is used to solve new problems through adaptation of results obtained from solving previously met problems. A new problem is

solved through finding a similar case in the records and through application of the solution related to that found case. An important feature of CBR is the ability to learn through collecting solutions of past problems and sharing them for solving new problems in future [6].

The proposed solution supports decision processes:

1. Choice of a supplier in the following cases: a necessity to change the supplier of a given material/service, a necessity to outsource some actions to appropriate suppliers in order to execute an order placed by a customer, at the stage of design and choices made by constructors of components and elements.
2. Estimation of the product price, based on an enquiry made by a customer.

In both cases a hybrid reasoning system was used, based on a case database and an expert system. An expert system supports the CBR at the stage of adaptation of found cases. A Rule-based Expert System (RES) is used to cooperate with small knowledge modules, which can be integrated in order to build a bigger database. Knowledge in the RES system is in the standard Horn clause format.

The mechanical engineering industry features a very vast group of suppliers from numerous branches, therefore a supplier assessment issue is so crucial for the final results of the execution of a taken order. Use of the suggested system allows for such a choice of a particular supplier of each component as to secure an appropriate level of supply made on time, to sustain a high level of quality of the ready product, as well as to keep the number of claims from the customers to the minimum. The system can support actions of constructors, engineers, purchase department staff and employees who prepare offers for customers.

As a result of a research conducted among mechanical engineering enterprises, two categories of actions taken during preparation of an offer for the customer based on their query can be distinguished. Those actions include an assessment, decision-making and actions related to the transfer of information. An offer is communicated to a customer who can reject it, accept it or undertake negotiations related to the price. In case of negotiations customer's requirements may be lowered in order to decrease the price.

A Case Based Reasoning system is a tool with an ability to gather, use and diffuse gained experience (including tacit knowledge) of employees responsible for preparation of an initial offer for a customer. Use of an expert system reinforces and broadens possibilities in reasoning and recommendations for the decisions being taken. In that way, joint use of CBR and ES allows to adjust previously found solutions to selected cases and solve current issues.

2.2. Knowledge Object Management System

A Knowledge Object Management System (KOMS) allows to handle knowledge represented by standardized elementary objects. Those objects can be linked through superiority – inferiority relations and connection relations into any structures fit for the purpose.

A primary KOMS object integrates a homogeneous formal structure with an ability to register various types of information (numerical, textual, graphical and other, also more complex elements). The structure of those elementary objects of knowledge allows to assign various attributes to them: descriptive and interpretative, classifying and evaluative, verbal and numerical (such as e.g. an ordinal number, an update date, an information source, an importance score, a reliability score, etc.) in order to characterize factual information.

In a knowledge system of general application, such as KOMS, textual description is the basic representation. A natural language description is particularly useful for classification of qualitative information with poor structuring that reflect sequential thought models. However, descriptive language has limited application in representation of quantitative information and information about multidimensional (e.g. spatial) objects, as well as in representation of complex, non-sequential dependencies (e.g. networked relations).

A KOMS system has a possibility to supplement a textual description with any MS Windows object. It can be a graphics or sound object, a video clip, an MS Word text file, an MS Excel spreadsheet, etc. Particularly, it is possible to attach objects which exist within the rule-based RES expert system and the Interactive Pair-matching System (IPS).

An elementary knowledge object of the KOMS is recorded as one entry in a database which has the following structure:

1. Object type – chosen from user pre-defined types.
2. Label (title).
3. Basic textual description (content).
4. Any MS Office object – OLE object.
5. Numerical, textual attributes, dates.
6. Other information fields – pre-defined by the user.

Moreover, an entry includes additional fields related to the system management. It is also possible to add further fields if needed.

Entries with elementary knowledge objects can be linked to each other in superiority-inferiority relations or connection relations into a hierarchical or networked structure (a directed graph) [7]. Each entry can be linked to any number of superior, inferior and peer entries. It is also possible to establish relations which form a loop. It is possible to register free entries which have no relation to any other entries.

2.3. Risk management support tool in a knowledge management supporting system in a mechanical engineering enterprise

Execution of orders for contracts in the analysed mechanical engineering enterprises proves legitimacy of utilization of the Project Management Institute methodology. Implementation of the functions of the tool called the Knowledge Object Management System (KOMS) allows to use it as a tool to support risk management in a mechanical engineering enterprise. For that purpose, it is necessary to make some additional assumptions to the concept of KOMS:

- an ability to define object types with a strictly set structure in a minimal scope (i.e. a list of object parameters which cannot be then removed without an ability to change types of those parameters) with an ability to add more parameters,
- an ability to define object types with a limited number of parameters (strictly set) without an ability to add more parameters,
- arrangement of an explicit relationship between some objects (pertains to objects of the "risk factor assessment" type: each of those objects should be explicitly assigned to one and one "risk factor" type of object only), which means that once assigned assessment cannot be used for any other object,
- arrangement of an object type with one superiority relationship (risk factor – contract), i.e. an object type which can be in a superiority relationship with only one other object,
- an ability to clone objects without preserving their superiority-inferiority relationship to other objects,
- an ability to clone objects which meet established criteria or are chosen by an user.

A "risk factor" object as a structure in the database should at least comprise the following parameters (types of parameters given in parentheses):

- a name of risk factor (text),
- an assignment of a factor to a contract ("None" or an ID of a contract) (assignment of a particular contract is made during establishment of a relation between a risk factor and a contract),
- a risk factor verification attribute (to be chosen from a list: "before verification", "after verification", "after conclusion of the contract"),
- owner of the risk entry (text),
- a risk scope (to be chosen from a list: "Design", "Supply", "Production", "Distribution", "Contract portfolio") as a parameter to classify the risk factors in the system,

- description of a risk factor (text),
- relationships with the list of risk assessment objects (including the probability and results in both quantitative and descriptive formats),
- result-limiting actions (as a text list).

The KOMS tool will allow to realize the following functions in relation to risk management support in a mechanical engineering enterprise:

- collection of knowledge on risk factors assigned to particular contracts,
- qualitative analysis of risk factors assigned to particular contracts,
- analysis of information on the state of assessment of risk factors at the early identification stage, at the stage of verification after preventive measures were proposed and at the stage of contract execution,
- collection of experiences from risk assessment as a result of comparative analysis of the state of risk factors during execution of a contract and after it was finalized,
- searching the database of risk factors according to a complex lookup key (with several conditions set, e.g. a particular value of the risk level, a particular contract or risk scope).

2.4. A tool to support human resource management processes in knowledge management system in mechanical engineering enterprise

Suggested solutions within a management support system in the range of knowledge management, which support resource management processes are primarily aimed at improvement of effectiveness and execution of contracts for particular orders of customers of the mechanical engineering industry. Support of management of human resources within the system comprises 4 main functionalities:

- employee database,
- external expert database,
- employee competence matrix,
- allocation of human resources to contracts.

The functionality of the “employee database” will include the following:

- entering a new employee,
- removing an employee,
- modifying information about an employee,
- searching for contracts/contract tasks executed by an assigned employee,
- searching for employees who have a required level of competence within the scope of a given task/process,
- searching for employees with particular experience in contract execution.

The functionality of the “external expert database” will include the following:

- entering a new expert,
- modifying information about an expert,
- searching for contracts/contract tasks executed by an assigned expert,
- searching for experts who have a required level of competence within the scope of a given task/process,
- searching for experts with particular experience in contract execution.

The employee competence matrix as a functionality in the designed system is a tool which allows to assign a certain level of competence (or a skill) to each employee of a mechanical engineering enterprise for a given business process (a task, an operation or type of skill). The designed system allows for a context-sensitive search of competence levels (and their modification) for the employees selected according to a custom lookup key. “Employee competence matrix” is a functionality related to the “allocation of human resources to contracts” functionality. Ability to edit the level of employee competence is limited to the group of employees who have a particular access level.

The functionality of the “allocation of human resources to contracts” will include the following:

- choice of a project team for execution of a contract,
- access control for contract execution tasks pertaining the the scope of human resources,
- introduction of changes in assignments of contract tasks,
- assessment of employees in executed contracts,
- a platform for co-operation between employees involved in execution of a given contract.

2.5. Rule-based expert system (RES)

A rule-based expert system (RES) was designed to co-operate with knowledge modules which can be integrated in order to expand the knowledge database. Knowledge in the RES system is kept in standard Horn clauses. Those are rules (expert implications) with one conclusion and with their conditional parts being conjunction statements. The conclusion is always a non-negative statement. Conditions and conclusions in those rules are logical sentences (not predicates), therefore they have no variables. That approach simplifies the architecture of an expert system and facilitates stability of its functioning. The conclusion process does not suffer from combinatorial explosions and looping, provided its rules are not internally contradictory.

Statements in the conditional parts of rules can be both simple (without negation) and negated, forming an extended knowledge database [8]. In that way,

knowledge can be represented by multiple levels of nested rules (statements in a simple form) and exceptions (negated statements). It allows for complex (non-monotonic) logical dependencies in a much simplified way.

2.6. Interactive Pair-Matching System

Interactive Pair-matching System (IPS) supports sequencing of objects on the basis of qualitative assessment. Comparison of two objects does not require to hold information about all other objects in the system memory. The standard method of paired comparisons has a disadvantage as it is time-consuming and requires matching all pairs for comparisons. Therefore, an original method of interactive pair comparisons was used in the IPS system. As compared to the standard method, this approach significantly decreases the required assessment time.

The original pair-matching method is based on a relation between the comparative score with database sorting. In that way, not all but only some pairs are matched for comparison. The choice of pairs and sequence of comparisons are a result of the work of a sorting algorithm and depend on the results of previously made comparisons.

Many sorting methods exist. In case of typical IT applications, effectiveness of an algorithm depends mainly on the number of operations such as data moving and control [9], whereas in case of an expert method of pair-matching the basic criterion of optimality is the number of matches. A method to minimize the number of comparisons was described by L. Ford and S. Johnson [10] and is called merge sorting. When compared to the standard method of pair-matching, a significant decrease in the number of comparisons is obtained, especially in case of databases with many objects (entries).

The number of matches can be decreased even furthermore in the following cases:

- a set of objects is initially ordered. Such situation occurs, for example, when the order of objects in a set is determined by a number of subsequent experts.
- Pairs or groups of objects in a set have the same assessment due to an assumed ordering criterion.

The system of interactive pair-matching uses a modified version of the Ford-Johnson method which facilitate a decrease of the average comparison operations without a risk of higher probability of negative results.

Execution of the pair-matching procedure makes the sorting algorithms choose subsequent pairs of objects. An expert may choose a factual answer or a control command for each of those pairs:

- Object A is more important than B.
- Object B is more important than A.

- Objects A and B are equally important.
- Undo – that command cancels the last assessment of an object pair. It is possible to undo an unlimited number of comparisons.
- End – stops the pair-matching session. Starting the session again allows to continue where it ended.

After all pairs chosen by the algorithm are assessed, an ordered set of objects is saved.

3. The scope of functionality of the it tools in the management supporting system in the range of knowledge management in mechanical engineering enterprises

Establishment of necessary functionality of the IT system supporting management in the range of knowledge management was based on the analysis of processes which are included in the knowledge management in a mechanical engineering enterprise. The structure of the system comprises the following modules:

- design,
- supply,
- production,
- distribution,
- supporting processes,
- system configuration.

The presented structure of the system focuses on supporting the decision-making processes executed as parts of contracts. It is necessitated by a need to provide vast support of knowledge processes in preparation, execution and control of contracts in order to enable their effective conclusion. Table 1 presents a set of tools to support knowledge management processes and their assignment to the main processes which manage the contract portfolio in a mechanical engineering enterprise.

4. Conclusions

1. The concept of a system supporting management in the range of knowledge management in mechanical engineering enterprises was designed on the basis of survey research. The model which served as the base for designing the IT system includes, as its key aspect, the current deficit of knowledge in the field of IT tools which support knowledge management in the key processes that manage the contract portfolio.

2. The designed system supporting management in the range of knowledge management in mechanical engineering enterprises included implementation of the following IT tools:
   - Case Based Reasoning (CBR),
   - Knowledge Object Management System (KOMS),
   - Risk Management Support Tool (RMST),
   - Human Resources Support Tool (HRST),
   - Rule-based Expert System (RES),
   - Interactive Pair-matching System (IPS).
3. The IT system which facilitates the suggested information tools will be a solution to support knowledge management aimed at increasing the effectiveness of management in the range of planning and supervision of the contract portfolio in a mechanical engineering enterprise.

Table 1. Assignment of tools supporting knowledge processes in management of the contract portfolio in a mechanical engineering enterprise in the Management Support System in the range of Knowledge Management

Customer portfolio management processes	**Knowledge process**			
	Knowledge collection	**Knowledge gathering**	**Knowledge transfer**	**Knowledge use**
Design	KOMS (record of good and bad practices) ES (collection and codification of knowledge in the form of expert system rules) CBR (codification of experiences in the form of cases)	KOMS (gathering of experience in the hyper-text form) CBR (gathering of knowledge in the form of verified cases)	KOMS (update and sharing of knowledge) RM (update and sharing of risk-related knowledge) CBR (possibility to share knowledge through access to the case database)	KOMS (use of electronic procedures for monitoring of business processes) ES (initial analysis of technical and organisational conditions of contract execution) ES (initial legal analysis of a contract) CBR (simplification of the contract price estimation process)
Supply	KOMS (record of good and bad practices, codification of procedures) ES (collection and codification of procedures) CBR (codification of experience in a form of cases)	RM (gathering of risk-related knowledge) KOMS (gathering of experience in the hyper-text form) CBR (gathering of knowledge in the form of verified cases)	RM (update and sharing of risk-related knowledge to system users) KOMS (update and sharing of knowledge to system users) CBR (possibility to share knowledge through access to the case database)	SE (initial selection and assessment of suppliers) KOMS (use of electronic procedures for monitoring of business processes) IPS (support of expert evaluation in the field of supply) CBR (support for constructors, employees of the purchase department, ensuring high quality of supply terms)
Production	KOMS (record of good and bad practices, codification of procedures) ES (collection and codification of procedures used as expert system rules)	RM (gathering of risk-related knowledge within the field of production) KOMS (gathering of experience and description of good practices in the hyper-text form)	RM (update and sharing of risk-related knowledge to system users) KOMS (update and sharing of knowledge to system users)	KOMS (use of experience, good and bad practices) ES, KOMS (use of electronic procedures for monitoring of business processes) ES (initial analysis of resource availability) IPS (support for expert evaluation)

Table 1. Continued

Customer portfolio management processes	Knowledge process			
	Knowledge collection	**Knowledge gathering**	**Knowledge transfer**	**Knowledge use**
Distribution	KOMS (record of good and bad practices, codification of procedures) ES (collection and codification of procedures used as expert system rules)	RM (gathering of risk-related knowledge within the field of distribution) KOMS (gathering of experience in the hypertext form) KOMS (gathering of good practices in the field of services)	RM (update and sharing of risk-related knowledge to system users) KOMS (update and sharing of knowledge to system users) KOMS (sharing of good practices in the field of services)	KOMS (analysis of customer claims) ES, KOMS (use of electronic procedures for monitoring of business processes) ES (support of analysis of possible decisions in the field of distribution) IPS (support of expert evaluation) KOMS (use of good and bad practices for the purposes of the service department)
Additional processes	KOMS (record of good and bad practices used in additional processes, codification of procedures) ES (collection and codification of expert knowledge used as expert system rules)	HRM (gathering knowledge about employees and external experts) RM (gathering of risk-related knowledge within the field of additional processes)	RM (update and sharing of risk-related knowledge to system users)	KOMS, HRM (location of knowledge resources, searching information about experts and their assessments) RM (risk minotioring within the field of additional processes) HRM (choice of project staff for execution of contracts) HRM (choice of project manager for execution of contracts)

REFERENCES

[1] Dohn K., Gumiński A., Zoleński W., (2011) "Assumptions for the creation of a system supporting knowledge management in an enterprise of mechanical engineering industry". Information systems in management XIII. Business intelligence and knowledge. pp. 19-27. WULS Press Warsaw.

[2] Dohn K., Gumiński A., Zoleński W. (2012) „Uwarunkowania przygotowania do implementacji systemu informatycznego wspomagającego zarządzanie wiedzą w przedsiębiorstwie produkcyjnym". Wiadomości Górnicze, nr 5/2012, str. 288-292.

[3] Jashapara A. (2006) Knowledge management. PWE, Warszawa.

[4] Kisielnicki J. (2008) Management Information Systems. Wydawnictwo Placet, Warszawa.

[5] Kowalczyk A., Nogalski B. (2007) Knowledge management. Conception and tools. Diffin. Warszawa.

[6] Witkowski T., (2000) Decyzje w zarządzaniu przedsiębiorstwem. Wydawnictwa Naukowo-Techniczne, Warszawa.

[7] Cormen T.H., Leiserson C.E., Rivest R.L:, (2007) Wprowadzenie do algorytmów. Wydawnictwa Naukowo-Techniczne, Warszawa.

[8] Niederliński A., (2000), Regułowe systemy ekspertowe. Wydawnictwo Pracowni Komputerowej Jacka Skalmierskiego, Gliwice.

[9] Wirth N. (2004), Algorytmy + struktury danych = programy Wydawnictwa Naukowo-Techniczne, Warszawa.

[10] Knuth D.E. (2002), Sztuka programowania. T. 3., Sortowanie i wyszukiwanie, s. 194, Wydawnictwa Naukowo-Techniczne, Warszawa.

The publication is financed from public science funds in the years 2010-2013 as the research project No. 03-0112-10 /2010 dated 09.12.2010.

INFORMATION
SYSTEMS IN
MANAGEMENT

Information Systems in Management (2014) Vol. 3 (1) 16–26

THE SYSTEM OF PLANNING AND MANAGEMENT OF THE MEDICAL SERVICES

DARIUSZ DOLIWA, WOJCIECH HORZELSKI, MARIUSZ JAROCKI

Faculty of Mathematics and Computer Science, University of Lodz

The system presented in this paper is designed to manage medical services at the Center for Rehabilitation and Cosmetology operating at the University of Computer Science and Skills in Lodz. The Center provides rehabilitation services and medical cosmetology. The purpose of the system is to create and modify the schedule of services provided by the Centre, including the availability of the resources, the preferences provided by the patient and the conditions resulting from the specific nature of these services. The system consists of the following components: a repository of resources, the databases describing the current schedule, the language of description of the relationships between the resources, a query language, with the editor supporting the query creation, and the two applications, the first generates a sequence of services required by the patient and the second modifies the current schedule to take account of the requested services.

Keywords: task scheduling, medical services managing, component oriented architecture

1. Introduction

There are many papers dealing with the topic of composition independent services based on semantic descriptions these services. Particularly noteworthy are the works [1-3]. The paper [4] provides an overview of the methodology of solving such problems. Some of these works consider static approaches, where flows are given as a part of the input, while the others deal with dynamically created flows.

One of the most active research areas is a group of methods referred to as AI Planning [5]. Several approaches use Planning Domain Definition Language [6]. The most complete specifications of an automatic composition system was described in [7]. The author proposed a solution based on a multi-phased composition using a uniform semantic description of services. Our approach uses similar methods. For service description was adopted language presented in papers [8-9]. The solution have been moved from the area of network services and implemented in a component-based business system.

2. System description

The system is designed to manage medical services at the Center for Rehabilitation and Cosmetology operating at the University of Computer Science and Skills in Lodz. The Center provides rehabilitation services in the form of medical visits, rehabilitation treatments and laboratory tests. This activity is not confined to individual patients, the majority of treatments are carried out to independent users of both individuals and institutions; the Center is also a service integrator for the external entities (branch offices, independent companies which want their services to be offered by the Centre). The purpose of the system is to create and modify the schedule of services provided by the Centre, including the availability of the resources, the preferences provided by the patient and the conditions resulting from the specific nature of these services. The planning involves searching for and allocation of the resources that meet the required needs. Designed for Center planning system should have the following characteristics:

- Users should be able to easily use the application provided by the Centre, have the ability to easily formulate their needs in a language close to natural (or by a simple graphical interface.)
- Resource allocation should be easily modified by employees without qualifications in the field of computer science.
- The system should be able to implement medical indications given as the result which is to be achieved (eg. a specific series of rehabilitation treatments of a particular type) rather than as a precise plan indicating where and by whom this goal should be pursued. Identifing the steps to achieve the goal should be automated by the system.
- Individual steps forming a plan can be implemented by a variety of the "service providers" (CIRK Center, cooperating entities etc.). The provided services may have the same functionality, but can vary according to the quality indicators (such as location, time, cost of the service). The System should allow to assess

the resulting plan and make adjustments by the operator (the resulting plan may be considered as a starting point for further manual adjustments that may be necessary during its implementation).

3. System architecture

The system consists of the following components (Figure 1.): a repository of resources, the database containing information about patients, doctors and physiotherapists and storing the current and past service reservations, the language of description of the relationships between the resources, the query language (with the editor supporting the query creation), and the two applications, the first generates a sequence of services required by the patient and the second modifies the current schedule to take account of the requested services.

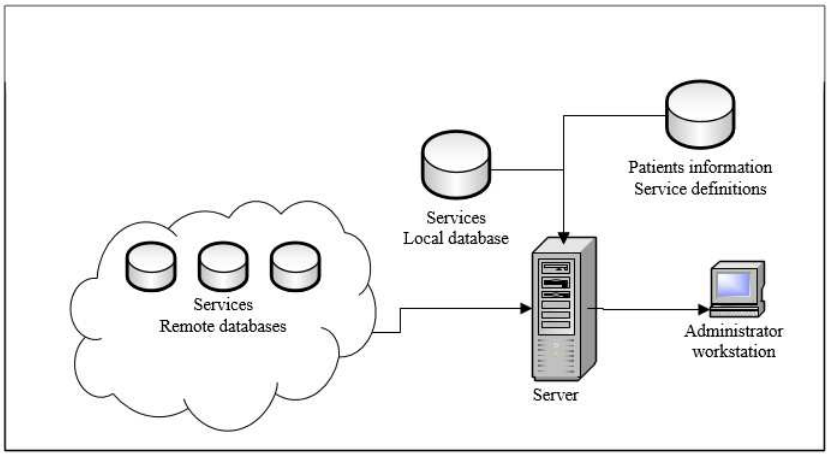

Figure 1. System architecture

A repository of the services for the system can be either a local database or the services obtained through the interface adapting the network services of the external providers. The repository allows you to query a list of the available treatments that meet the specified constraints, and returns the proposals of the service. The communication allows to reserve a service, in addition it is also possible to cancel a service (a patient gave up a treatment).

The language created for the application allows you to define the relationship between the objects available in the system (such as a patient, a treatment, a laboratory test, a physical therapist, a physician) and the services (a patient's registration, a medical examination, a laboratory test, a specified type treatment). It

also allows you to define the preconditions that must be fulfilled for the service to be performed, indicates how the objects participating in the service will be modified by the service and it also allows you to define the sequence in which the individual services will be performed.

The system contains also a query language that allows the patient to determine the required services and the conditions that they have to meet (eg. a location, the time of the service, a request for a specific doctor or a physiotherapist). Since the questions are to be constructed by the user, in the system there is available the query generator, which through a simple graphical interface allows the user to express his request in the system language.

After entering the user's query the application is run and generates a sequence of (or a set of the sequences) services that meet the needs of the patient and fill described by the language relationships between the offered services. Then there is verified their availability in the repository where the relevant services are reserved or we get the information that posed requests can not be processed. If the request can be made, based on the received plan, the current schedule of the tasks to be implemented by the Centre is modified.

4. System services

The application is designed to allow the end user (the patient) to schedule a visit to the Centre offering a wide range of medical services. It may be a visit to a general practitioner, a specific type of specialist doctor, execution of a single medical test or a group of tests, and finally planning one or more rehabilitation treatments.

List of available medical specialties, types of tests and types of treatments and their characteristics are defined by the service provider in the database. In the system we have objects, such as *Patient*, *Doctor*, *Physiotherapist*, *Test* and the services provided to the patient. These services are divided into types as a registering in the Centre (*Registration*), a medical examination (*VisitDoctor*), a laboratory test (*Therapy*) and a rehabilitation treatment (*Treatment*). Each of these services has some properties specified on the server (the service definitions). Each may have some preconditions for running and can cause changes in the system by creating new or changing the existing objects. The services may also leave during runtime some dynamic information about their use (*Trace* of service) allowing to save the information about the number of positions within a service sequence, time of the service execution, its cost, and other properties. In addition, this makes it possible to impose a particular sequence of services for the patient by the administrator.

An Object *Patient* represents the patient and has the following properties:

```
Patient := {id : int, name : last_name : string, first_name :
string,date_of_birth : date, pesel : string, phone : string,
sms : Boolean, diagnosis : string, symptoms : string }
```

It is created at the time of starting the use the system by the patient (via *Registration* service) and is a carrier of information about the patient and the history of its treatment.

Objects *Doctor* and *Physiotherapist* represent CIRK employees. Besides the basic information (personal information) they contain the data needed for planning the Center's work (such as an availability, a type of treatments carried out, etc.).

Objects *Test* and *Therapy* are similar in a structure and store information about tests performed in the Centre. They are created by the system administrator, if necessary - when the Centre plans to offer a new service - a test or a treatment. They have the following characteristics:

```
id     : int
name   : string
result : string
comment: string
```

The services offered in the system are administrative (*Registration*), or medical services (such *VisitDoctor*, *Test*, *Treatment*, etc..). A *Registration* service has no prerequisites and creates a new patient (an object *Patient*). It also sets the data of the patient's properties. It is a service that needs to be done once for each patient, before any other services. If the patient has already been created then this service will no longer be called by his subsequent visits to the clinic.
In the language of the system it can be described as follows:

```
Registration:
IN           :     null
OUT          :     p:Patient
PRECONDITION :     null
POSTCONDITION:     isSet(p.id) and isSet(p.name) and
                   isSet(p.last_name)and isSet(p.first_name)
                   and isSet(p.pesel)and isSet(p.phone)and
                   isSet(p.sms)
```

The other services offered by the system correspond to medical services. The objects of *VisitDoctor* type are services like *GeneralPhysician*, *Laryngologist*, *Cardiologist*, etc., representing a visit to a doctor of a particular specialty. Calling these services it requires the existence of a *Patient* object and sets the property diagnosis of this object.

The objects of *Laboratory* type are all kinds of laboratory tests (blood testing, DNA testing, etc.), and *Treatment* type are the treatments offered by the Centre (*Rehabilitation*, *Physiotherapy*, *Cytology* etc.).

The *VisitDoctor* and *Treatment* services have the following definitions:

```
IN                  :     p:Patient
OUT                 :     p:Patient and tr:TreatmentTrace
PRECONDITION        :     isSet(p.id)
POSTCONDITION       :     isSet(p.id) and
                          isSet(p.recommendation) and
                          isSet(tr.start) and isSet(tr.stop)
                          and isSet(tr.location)and
                          isSet(tr.servicename) and
                          isSet(tr.servicetype)
```

and they differed only by the values for the fields in the *servicetype* and *servicename*. The system assumes that the addition of a new medical specialty, new type of medical test or a new type of rehabilitation treatment requires the definition of the new services in the database on the server.

Planning a visit to the patient in the Centre will require the formulation of conditions of this visit in the system language (the query language), that determine the information held on the input (a requirement of the relevant objects and their properties) and the requirements to be met at the exit.

```
Input:
List of objects: . . .
Precondition: . . .

Output:
List of objects: . . .
Postcondition: . . .
```

An example scenario of a planned visit when the patient uses the system for the first time and would like to arrange a medical visit in order to determine the necessary treatments will look as follows:

```
Input:
List of objects:    null
Precondition:       null
Output:
List of objects:    p : Patient
Postcondition:      isSet(p.id) and isSet(p.name) and
                    isSet(p.last_name)and isSet(p.first_name)
                    and isSet(p.pesel) and isSet(p.phone) and
                    isSet(p.sms)and isSet(p.recommendation)
```

In this case, we do not require any object on the input, and we want to be created a new object representing the patient with properties that describe it and recommend further procedure. In more complex cases the query can be more complicated such as when a patient Kowalski (having already an account in the system and being assigned an identifier, has also staked a diagnosis) want to visit a dermatologist and perform blood test (the order of the sequence is important), in Lodz in the first week of January 2012. The query in this case looks as follows:

```
Input:
List of objects:    JKowalski:Patient,  T_GP:Trace
Precondition:       T_GP.service = ”General_physician”) and
                    isSet( JKowalski.diagnosis)
Output:
List of objects:    JKowalski:Patient,  T_GP: Trace; T_D1:
                    Trace, T_D2:Trace;
Postcondition:      (T_D1.servicename = ”Dermatologist”)  and
                    (T_D2.servicename=”BloodTest”) and
                    T_D1.level<T_D2.level and
                    month(T_D1.start)=”January” and
                    month(T_D2.start=”January” and
                    T_D.location=”Lodz”
```

In this example, we also apply the aforementioned object *Trace*, which allows you to control which services were developed earlier and for example determine the order of the services implementing the plan.

Writing queries in the system can be quite complicated for the person administering it (we assume that he is not a computer scientist). For this reason, in the system was created the query generator, which through a simple graphical user interface (Figure 2.) allows the user to express his request.

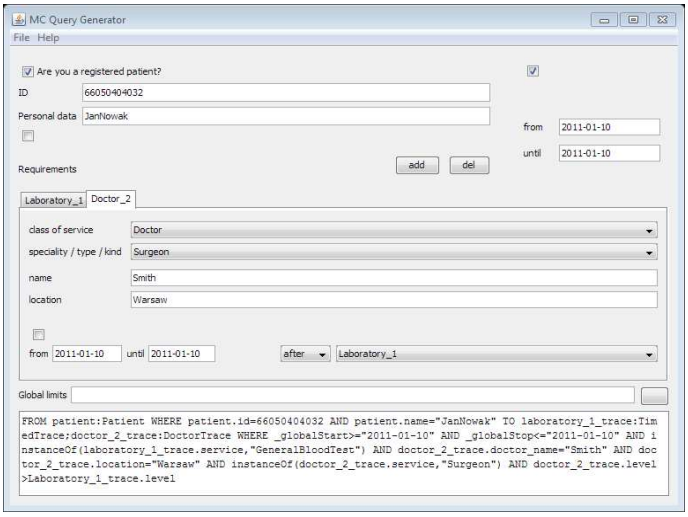

Figure 2. Query generator

The queries based on the service definitions will generate an initial plan which is a sequence of services implementing the patient's request. Then, for the resulting sequence must be associated with specific services (after checking their availability). For this purpose we need a repository of these services.

5. Repository of services

The services provided locally use the register in a database maintained on the server, but it is also possible for the patients of the Center to use the services from external sources which are provided by the web services. A record whether local or remote must allow for an inquiry (enables querying for available services without making commitments), a booking (filling in the patient for a specific treatment, carried out by a specific physiotherapist, in a given time and place) and cancellation (cancellation of a recording and release the associated resources).
A local database used in the Center has the structure shown in Figure Figure 3 and consists of tables such as Patients, Physioterapists, Doctors describing patients of the Center (the identity of the patient and contact details, symptoms and medical advice) and working in the Center doctors and physiotherapist (eg. personal data, represented specialties).

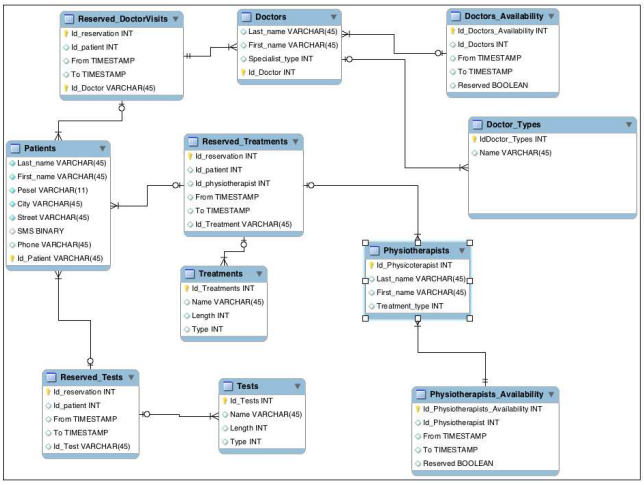

Figure 3. Database structure

The tables *Treatments, Tests* define the types of treatments and tests offered in the Center. Other tables are responsible for storing information about the availability of services and a staff, and are modified during creating a schedule of tasks. Additionally, the application, which generates a schedule based on these tables, selects the available services that meet patient's demand.

6. Plan generator

After defining a user's query the applications, which are a part of the system, create a plan that satisfies the conditions resulting from the query. The operations of generating the plan have two phases. In the first, we make a sequence of types of services that match the query, created on the defined dependencies between the services (eg preconditions services require the existence of objects created by other services) or defined directly by the user (through the references to objects *Trace*). For some more general questions we may have multiple correct sequences and then the user of the system is responsible for choosing one of them. In the second phase for the existing sequence the specific services are looking for in the repository (requests for a quotation mode), which is used by the system. They have already defined the properties such as a place of execution, a person responsible for their performance, the time in which they are available and the cost assigned to them . Including these characteristics and their compatibility with the query arises the actual plan of the services that can be offered to the patient. As in the first phase, if

there are several correct plans (corresponding to the query) the user is responsible for the selection. The selected services are marked as reserved in the repository and included in the schedule of tasks to be implemented by the Centre.

7. Final remarks

The system is a specialized implementation of the concept, which can be applied to various domains, enabling to build an integration system for distributed services of a common characteristic. More generally, a similar system can be implemented in every domain in which we have to plan an access to some resources. The system is based on concept of PlanICS [9], which was adapted to medical services modelling and scheduling.

A further contribution of the system is in an extended language of queries, enabling to express more requirements occurring is practice.

REFERENCES

[1] J. Rao. *Semantic Web Service Composition via Logic-Based ProgramSynthesis.* PhD thesis, Dept. of Comp. and Inf. Sci., Norwegian University of Science and Technology, 2004.

[2] J. Rao, P. Küngas, and M. Matskin. *Logic-based web services composition: From service description to process model.* In Proc. of the IEEE Int. Conf. on Web Services (ICWS'04). IEEE Computer Society, 2004.

[3] D. Redavid, L. Iannone, and T. Payne. *OWL-S atomic services composition with SWRL rules.* In Proc. of the 4th Italian Semantic Web Workshop: Semantic Web Applications and Perspectives (SWAP 2007), 2007.

[4] B. Srivastava and J. Koehler. *Web service composition - current solutions and open problems.* In Proc. of Int. Workshop on Planning for Web Services (at ICAPS 2003), 2003.

[5] M. Klusch, A. Gerber, and M. Schmidt. *Semantic web service composition planning with OWLS-XPlan*. In Proc. of the 1st Int. AAAI Fall Symposium on Agents and the Semantic Web, pages 55–62. AAAI Press, 2005.

[6] D. McDermott, M. Ghallab, A. Howe, C. Knoblock, A. Ram, M. Veloso, D. Weld, and D. Wilkins. *PDDL - the Planning Domain Definition Language - version 1.2.* Technical Report TR-98-003, Yale Center for Computational Vision and Control, 1998.

[7] S. Ambroszkiewicz. enTish: An Approach to service Description and Composition. ICS PAS, Ordona 21, 01-237 Warsaw, 2003.

[8] D. Doliwa, W. Horzelski, M. Jarocki, A. Niewiadomski, W. Penczek, A. Półrola, and M. Szreter: *Web Services Composition - From Ontology To Plan By Query*. Control & Cybernetics 40(2) pages 315-336, 2011.

[9] D. Doliwa, W. Horzelski, M. Jarocki, A. Niewiadomski, W. Penczek, A. Półrola, M. Szreter and A. Zbrzezny: PlanICS - *A Web Service Composition Toolset.* Fundamenta Informaticae 112(1), pages 47-71, IOS Press, 2011.

[10] D. Doliwa, W. Horzelski, M. Jarocki, A. Niewiadomski, W. Penczek, A. Półrola and J. Skaruz: *HarmonICS - a Tool for Composing Medical Services.* Postproc. of the 4th Central European Workshop on Services and their Composition (ZEUS'12), CEUR Workshop Proceedings vol.847, pages 25-33, 2012.

INFORMATION SYSTEMS IN MANAGEMENT

Information Systems in Management (2014) Vol. 3 (1) 27−39

CONSENSUS DETERMINING ALGORITHM FOR SUPPLY CHAIN MANAGEMENT SYSTEMS

MARCIN HERNES [a)], JADWIGA SOBIESKA-KARPIŃSKA[b)]

[a)] *Department of Economic Informatics, Wrocław University of Economics*
[b)] *Department of Economic Communication, Wrocław University of Economics*

The purpose of article is to elaborate a consensus determination algorithm in supply chain management support systems, which may lead to achieving a greater flexibility and effectiveness of such systems. Using consensus methods in resolving the conflict of knowledge, in other words, determining a variant to be then presented to the user, based on the variants proposed by the system, may lead to shortening the variant determination time and to reducing the risk of selecting the worst variant. As a consequence, supply chain management might become more dynamic, which obviously influences the effectiveness of the operation of particular organizations and the entire supply chain. The originality is using consensus method to resolve knowledge conflicts in SCM systems to help decision-maker to take decision earning satisfy benefits.

Keywords: Supply Chain Management Systems, knowledge conflicts, consensus methods

1. Introduction

Nowadays a supply chain constitutes the crucial component in the operation of enterprises in the turbulent economic situation. What determines competitive advantage to an increasingly greater extent, is not only the quality and price of a product but also efficient organisation of supplies of materials, raw materials and finished goods at the lowest costs and appropriate level of customer service en-

sured. This leads enterprises to seeking the best strategies permitting effective supply chain management [3, 8].

It has been noticeable in the recent years that enterprises have become more interested in systems oriented towards SCM (Supply Chain Management) [10, 11]. It is indicated more and more often in the relevant literature [12, 13] that such systems ought to dynamically respond to the market needs, which contributes to increasing the value of all enterprises participating in the execution of a supply chain. However, the situation where the SCM system presents diverse variants of product flow between individual co-operators to the user occurs frequently [16]. Each of these variants can be different, that is have different values of the attributes (characteristics) describing the product flow (for example, the delivery date can be an attribute). This means that there is a conflict of knowledge between the variants generated by the SCM system. The variant selection should bring satisfying benefits to the user, that is this ought to be a variant which allows the delivery of goods in an appropriate quantity and within an appropriate time at the lowest possible costs and risk. If the user has to make the decision on which of these variants to choose alone, this process is obviously time-consuming and involves the risk of selecting the wrong variant. For instance, one might choose the variant in which, despite the timely delivery and low costs, the size of the goods batch is too big and the enterprise will have to incur the expenses of warehousing the goods. One could also select the optimum variant in terms of the size of the batch and costs but with a considerable risk of delay in delivery. Therefore, choosing a wrong or high-risk variant can cause disturbances in the performance of production or service processes in the enterprise. This will undoubtedly influence the effectiveness of its functioning and even, in an extreme case where, for instance, the customer dissatisfaction with the price, quality and timeliness of the services provided by this enterprise increases, result in its bankruptcy. Due to such problems, the conflict of knowledge should not be resolved by a human but rather by the system automatically and on a real-time basis. Various methods of resolving such conflicts can be found in the relevant literature, e.g. negotiations [4], or deductive-computing methods [2]. Negotiations enable a good resolution to the conflict of knowledge by achieving a compromise but they require an exchange of a large number of messages between the system components, as a result of which the operation of the SCM system on a real-time basis often becomes difficult or even impossible. Deductive-computing methods, for instance ones based on the game theory (game theory is a discipline of science closely related to the problems of cooperation and conflict resolving in multiagent systems, it involves construction of mathematical models of conflicts and cooperation as elements of human interaction), classical mechanics (they are employed in multiagent systems that require cooperation between a large number of agents – hundreds or even thousands of agents – such as in systems designed to reach highly distributed and dynamic goals) or a multi-criteria

method (choice one of solutions on the basis of multiple criteria) permit the achievement of high computational capacity of the system but they do not ensure a correct resolution to the conflict of knowledge, because often choice one of solution is related is associated with a high level of risk.

For the purpose of eliminating the presented problems, it is possible to employ consensus methods, which allow resolving the conflict of knowledge at a time close to the real time [15] and at the same time ensure the achievement of a good compromise [9]. This is because in the case of a consensus each of the parties is taken into account, "loses" as little as possible and makes its contribution to the consensus, all parties accept the consensus, which means that the consensus represents all parties to the conflict. Consensus determination algorithms regarding various decision-making areas, such as weather forecasting, finances, or environmental monitoring, can be found in the relevant literature. However, no such consensus determination algorithms which could be used in support systems for supply chain management have been developed thus far.

Hence, the present paper aims at developing a consensus determination algorithm in supply chain management support systems, which may lead to achieving a greater flexibility and effectiveness of such systems. Article is a continuation of research presented in [16, 17]. The first part of article presents a short characteristics of consensus methods. Next, the distance functions between variants are presented. The algorithm of consensus determining was elaborated in the final part of article.

2. Consensus methods

The relevant literature [9] defines consensus as an agreement and originates from the choice theory. Consensus is determined based on the existing solutions to a given problem, is very close to them, but does not have to be one of these solutions. Hence, the supply chain management variant presented to the used does not need to be a variant determined by the SCM system. It can be a totally new variant formed on the basis of the existing variants (that is ones determined by the SCM system). Owing to that all variants generated by the SCM system can be taken into consideration. It needs to be noticed that since supply chain management is a continuous process, the selection from among the variants generated by the system is also made in a permanent manner. The time allocated for making this choice is very short because the system may generate new variants of solutions after a moment and a new choice needs to be made. So if the SCM system generates a few or ten-odd solution variants, one new variant can be determined automatically based on these variants and next presented to the user. Owing to that all variants generated by the SCM system can be taken into account. Such a behaviour permits, among others, shortening the target variant determination time. The user does not have to

analyse individual variants and contemplate their selection as the system will perform these actions automatically (of course user always have a possibility to analyse individual variants manually). It needs to be emphasised that due to the continuity of the supply chain management process, it is often impossible for the user to analyse the variants. This is because the time between the successive choices is so short that one is unable to make a decision without aid from the system. Employing consensus methods also allows reducing the risk of selecting the worst variant since all variants are taken into consideration in the case of a consensus. It should be noticed that this risk is extremely high because selecting a wrong variant may result, for instance, in untimely deliveries or their high cost intensity, which can lead to disturbances in the goods production or service provision processes and, in consequence, to reduced economic effectiveness of an enterprise or even its bankruptcy. Applying consensus methods, in turn, may result in a more flexible, effective and less risky execution of the supply chain management process.

The research conducted by the authors of the paper [16] show that the SCM system modules related to suppliers, producers, wholesalers, retailers and individual customers – based on the information obtained from transactional and analytical systems – due to various criteria or information analysis methods (such as the lowest price, the shortest lead time, non-linear programming, genetic algorithms, multi-criteria methods) generate diverse solution variants for individual supply chain components [14]. These variants differ in terms of attributes and values of these attributes. Hence, a conflict between these variants emerges in the system. Therefore, the user needs to choose a variant to be implemented at a given time. Such a choice in not easy because one is never sure which of the variants may bring results that will be satisfying to the user. This is why such a choice involves a high risk, and at the same time the user is incapable of analysing the considered variants due to time limits arising from the fact that decisions need to be made very quickly for supply chain management to be effective. Hence, conflicts of knowledge occur when the parties to the conflict assign different values to the same objects in the world and the same characteristics [16]. If the SCM system generates different solution variants (for example due to the employment of diverse support methods for supply chain management), the conflict of knowledge may refer to a characteristic such as “quantity” (in the generated variants the quantity of the goods to be shipped can differ), “cost” (in the generated variants delivery costs can differ), or “time” (in the generated variants delivery can be performed at different hours).

Resolving this type of conflicts is extremely important because only then is there a possibility to determine the right variant for the user, one being a compromise between the variants generated by the SCM system. Owing to that, each time the system generates different variant versions, the user obtains satisfying solutions characterised by a low risk level, which means that they will be released form the necessity to make a choice, which, as was stated above, is frequently impossible

due to time limits. As a consequence, this may contribute to timeliness, appropriate volume of the supply batches and low cost level. If the system ignores this aspect, the user might have problems with proper supply chain management. Selecting a wrong or high-risk variant can result in problems with maintaining production continuity, increased warehousing costs, or lack of financial liquidity and, as a result, decreased economic effectiveness of the enterprise. Therefore, the use of consensus methods will permit resolving conflicts of knowledge and, as a consequence, streamline supply chain management.

Determining consensus consist of three major stages (Fig. 1). In the first stage it is necessary to carefully examine the structure of the set of all the variants generated by the system SCM, or specify characteristics that represent these variants and the domains of its values (this aspect was presented at [16]). Structures of variants are knowledge structures of SCM. In the second stage it is necessary to define the distance functions among particular variants (it was presented at [17]) and to define set of variants (a profile), on the basis of which consensus will be determine. The third stage is an elaborate of consensus determining algorithms - the determining of such a variant, that the distance between this variant (consensus), and the individual variants generated by SCM is minimal (according different criterions). So, the consensus is not the average.

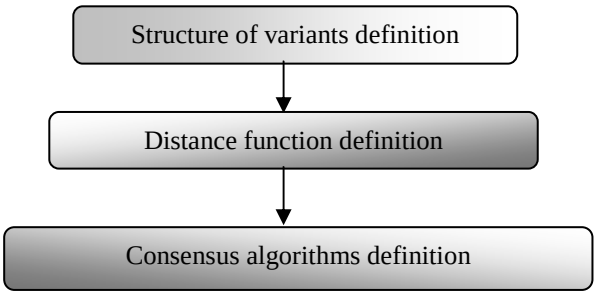

Figure 1. The stages of consensus determining.
Source: own preparation on the basis of [16]

In order to elaborate the algorithms of consensus dethermining it can be use a structure defined at the paper [16] in the following way (the first stage of consensus determining):

Definition 1
Let:
Set of products $T = \{t_1, t_2, \ldots, t_N\}$
Set of places $M = \{m_1, m_2, \ldots, m_L\}$

The structure of variant is called the following sequence:

$$W = \left\{ \left\langle t_1, m_{p1}, m_{q1}, dt_{m_{p1}}, dt_{m_{q1}}, i_1, k_1 \right\rangle, \left\langle t_2, m_{r2}, m_{s2}, dt_{m_{r2}}, dt_{m_{s2}}, i_2, k_2 \right\rangle, ..., \left\langle t_N, m_{xN}, m_{yN}, dt_{m_{xN}}, dt_{m_{yN}}, i_N, k_N \right\rangle \right\}$$

where:

$p, q, r, s, x, y = \{1..L\}$,

$dt_{m_{p1}}, dt_{m_{r1}}, ..., dt_{m_{xN}}$ - date and time of sending of product $t_1, t_2, ..., t_n$ from the place $m_p, m_r, ..., m_x$,

$dt_{m_{q1}}, dt_{m_{s1}}, ..., dt_{m_{yN}}$ - date and time of receiving of product $t_1, t_2, ..., t_n$ at the place $m_q, m_s, ..., m_y$,

$i_1, i_2, ..., i_N$ - the amount of transported products $t_1, t_2, ..., t_n$ (the size of the batch),

$k_1, k_2, ..., k_N$ - the cost of transport $t_1, t_2, ..., t_n$.

This definition allows the representation of individual variants of solutions in the form of uniform structure. It is complex, multi-value structure consist of different types of data.

3. Distance functions

The paper [17] suggest the distance function between structures of variants. It is the second stage of consensus determining. It must be noted that calculation of a distance between two variant structures may be based on calculation and summation of distances between individual elements of those structures.

For the purpose of defining time distance between two dates, let us assume that chronon, that is the smallest unit of time [1, 5] equals one minute (this degree of accuracy seems sufficient, since transportation of goods in practical application cannot be accomplished with down-to-one-second accuracy). Naturally, this assumption does not preclude one from adopting other time units as chronons. Therefore, definition of this function is as follow:

Definition 2

Distance ϑ between two dates $dt1$ and $dt2$ in the structure of variants is called the function:

$$\vartheta(dt1, dt2) = \left| dt1 - dt2 \right|.$$

The example is distance between dates: 10-11-2012 15:00 and 11-11-2012 16:30, which equal 1 day, 1 hour and 30 minutes that is 24 * 60 + 90 =1530 minutes.

In considering the distance between number of the product and the costs of transport, it can be use the function used in many papers [e.g. 6, 7, 9] specifying the distance between real numbers:

Definition 3

The distance between numbers *x, y* belonging to the string composed with *m* real numbers is called follow function*:*

$$\chi(x,y)=\frac{1}{m}|x-y|.$$

The following example illustrates this definition.

Let's $m=3$ and string of numbers is following: {2, 4, 8}. The distance between numbers 2 and 4 equals $=\frac{1}{3}|2-4|=\frac{2}{3}$, the distance between numbers 2 and 8 equals $=\frac{1}{3}|2-8|=2$, whereas the distance between numbers 4 and 8 equals $=\frac{1}{3}|4-8|=1\frac{1}{3}$.

At the article [17] the distance between two variants is defined as follow:

Definition 4

The distance Ψ between two structures of variants:

$$W^{(1)}=\left\{\left\langle t_1^{(1)}, m_{p1}^{(1)}, m_{q1}^{(1)}, dt_{m_{p1}}^{(1)}, dt_{m_{q1}}^{(1)}, i_1^{(1)}, k_1^{(1)}\right\rangle,\ldots,\left\langle t_N^{(1)}, m_{xN}^{(1)}, m_{yN}^{(1)}, dt_{m_{xN}}^{(1)}, dt_{m_{yN}}^{(1)}, i_N^{(1)}, k_N^{(1)}\right\rangle\right\}$$

$$W^{(2)}=\left\{\left\langle t_1^{(2)}, m_{p1}^{(2)}, m_{q1}^{(2)}, dt_{m_{p1}}^{(2)}, dt_{m_{q1}}^{(2)}, i_1^{(2)}, k_1^{(2)}\right\rangle,\ldots,\left\langle t_N^{(2)}, m_{x1}^{(2)}, m_{y1}^{(2)}, dt_{m_{x1}}^{(2)}, dt_{m_{y1}}^{(2)}, i_N^{(2)}, k_N^{(2)}\right\rangle\right\}$$

is called following function:

$$\Psi(W^{(1)},W^{(2)})=\sum_{j=1}^{N}\vartheta(dt_{m_{pj}}^{(1)},dt_{m_{pj}}^{(2)})+\vartheta(dt_{m_{qj}}^{(1)},dt_{m_{qj}}^{(2)})+\chi(i_j^{(1)},i_j^{(2)})+\chi(k_1^{(1)},k_1^{(2)})$$

Presented definition enables to calculate the distance between the two structures of variants. However, in order to calculate the distance between one structure (for example consensus), and more other structures (for example the variants generated by system), it should be proceed in the following way:

- calculate the distance between considered structure and each of the other individual structures,
- calculate the sum of these distances.

The algorithm of consensus determining (the third stage of consensus determining), using presented distance function, is elaborated in the next part of article.

4. Consensus determining algorithm

The postulated method of distance calculation may be employed in design of consensus algorithms. The consensus is determining on the basis of set of variants generated by system, called the profile, defined as follow:

Definition 5

The profile $A = \{W^{(1)}, W^{(2)}, ..., W^{(R)}\}$ is called set M variants, such that:

$$W^{(1)} = \left\{\left\langle t_1^{(1)}, m_{p1}^{(1)}, m_{q1}^{(1)}, dt_{m_{p1}}^{(1)}, dt_{m_{q1}}^{(1)}, i_1^{(1)}, k_1^{(1)}\right\rangle, ..., \left\langle t_N^{(1)}, m_{xN}^{(1)}, m_{yN}^{(1)}, dt_{m_{xN}}^{(1)}, dt_{m_{yN}}^{(1)}, i_N^{(1)}, k_N^{(1)}\right\rangle\right\}$$

$$W^{(2)} = \left\{\left\langle t_1^{(2)}, m_{p1}^{(2)}, m_{q1}^{(2)}, dt_{m_{p1}}^{(2)}, dt_{m_{q1}}^{(2)}, i_1^{(2)}, k_1^{(2)}\right\rangle, ..., \left\langle t_N^{(2)}, m_{x1}^{(2)}, m_{y1}^{(2)}, dt_{m_{x1}}^{(2)}, dt_{m_{y1}}^{(2)}, i_N^{(2)}, k_N^{(2)}\right\rangle\right\}$$

$$\vdots$$

$$W^{(R)} = \left\{\left\langle t_1^{(R)}, m_{p1}^{(R)}, m_{q1}^{(R)}, dt_{m_{p1}}^{(R)}, dt_{m_{q1}}^{(R)}, i_1^{(R)}, k_1^{(R)}\right\rangle, ..., \left\langle t_N^{(R)}, m_{x1}^{(R)}, m_{y1}^{(R)}, dt_{m_{x1}}^{(R)}, dt_{m_{y1}}^{(R)}, i_N^{(R)}, k_N^{(R)}\right\rangle\right\}$$

In order to determine consensus of defined profile it is necessary to elaborate theorem, on the basis of which specified will then be consensus algorithm.

Theorem 1

Let $pr(dt_{xy}), pr(i_y), pr(k_y)$ denote respectively ascending order of values $dt_{xy}^{\ L}, pr_y^{\ L}, k_y^{\ L}$ $(L = 1, ..., R)$,

$l^1_{dt_{xy}}$ denote $(R+1)/2$ element $pr(dt_{xy})$,

$l^2_{dt_{xy}}$ denote $(R+2)/2$ element $pr(dt_{xy})$,

$l^1_{i_y}$ denote $(R+1)/2$ element $pr(i_y)$,

$l^2_{i_y}$ denote $(R+2)/2$ element $pr(i_y)$,

$l^1_{k_y}$ denote $(R+1)/2$ element $pr(k_y)$,

$l^2_{k_y}$ denote $(R+2)/2$ element $pr(k_y)$ and

$$CON = \left\{ \left\langle CON(t_1), CON(_{p1}), CON(m_{q1}), CON(dt_{m_{p1}}), CON(dt_{m_{q1}}), CON(i_1), CON(k_1) \right\rangle, ..., \right.$$
$$\left. \left\langle CON(t_N), CON(m_{xN}), CON(m_{yN}), CON(dt_{m_{xN}}), CON(dt_{m_{yN}}), CON(i_N), CON(k_N) \right\rangle \right\}$$

will be the consensus according given profile.

then for each dt_{xy}

$$l^1_{dt_{xy}} \le dt_{xy}{}^L \le l^2_{dt_{xy}} \Rightarrow dt_{xy}{}^L \in CON(dt_{xy}),$$
$$(l^1_{dt_{xy}} > dt_{xy}{}^L) \vee (dt_{xy}{}^L l^2_{dt_{xy}}) \Rightarrow dt_{xy}{}^L \notin CON(dt_{xy}),$$

and
for each $i_y{}^L$

$$l^1_{i_y} \le i_y{}^L \le l^2_{i_y} \Rightarrow i_y{}^L \in CON(i_y),$$
$$(l^1_{i_y} > i_y{}^L) \vee (i_y{}^L l^2_{i_y}) \Rightarrow i_y{}^L \notin CON(i_y)$$

and
for each $k_y{}^L$

$$l^1_{k_y} \le k_y{}^L \le l^2_{k_y} \Rightarrow k_y{}^L \in CON(k_y),$$
$$(l^1_{k_y} > k_y{}^L) \vee (k_y{}^L l^2_{k_y}) \Rightarrow k_y{}^L \notin CON(k_y).$$

Proof 1

The ascending order can be presented as points on the axis of Reals and then the shortest distance to all points of a given range is a point lying in the middle of this range.

On the basis of theorem 1 it can be elaborate algorithm which determine consensus of profile of variants. Algorithm function in this way, that ascending order of values dt_{xy} is determining with all the variants and so it is with the values i_y and k_y. Then be calculated between which values in these systems must be the value which is a consensus. In the next step, set the value of consensus of products and places in this way, that with profile it is selected values of these attributes of this variant, in which the distance between a cost of this variant, and cost of variant determining by consensus, is minimal. The algorithm terminates when all elements of variant are verified and consensus is determining. The formal definition is as follows:

Algorithm 1

Data: Profile $W=\{W^{(1)}, W^{(2)}, \ldots W^{(R)}\}$ consist of R variants.
Result: Consensus

$$CON=\left\{\left\langle CON(t_1), CON(_{p1}), CON(m_{q1}), CON(dt_{m_{p1}}), CON(dt_{m_{q1}}), CON(i_1), CON(k_1)\right\rangle,\ldots,\right.$$
$$\left.\left\langle CON(t_N), CON(m_{xN}), CON(m_{yN}), CON(dt_{m_{xN}}), CON(dt_{m_{yN}}), CON(i_N), CON(k_N)\right\rangle\right\}$$

according W.

START
Step 1: Let $CON(x)=0$.

Step 2: $j:=dt_{xy}$.
Step 3: Determining $pr(j)$.
Step 4: $l_i^1 = (R+1)/2$, $l_i^2 = (R+2)/2$.
Step 5: $l_j^1 \le CON(j) \le l_j^2$.
Step 6: If $j=dt_{xy}$ then $j:=i_y$. Go to: Step 3.
If $j:=i$ then $j:=k_y$. Go to: Step 3.
If $j:=k_y$ then STOP. Go to: Step 7.
Step 6: Determining $CON(t_x), CON(m_{px}), CON(m_{px})$, which meets the following conditions $\min(\chi(CON(k_x), k_x)^{(2)})$.
STOP

Computational complexity is $O(N^2R)$.

Using elaborated algorithm in SCM system allow skipping the analysis of individual variants made by decision maker. At the same time, it should be noted that, in order to determining the consensus, the variants of solutions must be represented in the form of uniform data structures in the individual elements of the supply chain (for example, the structure representing the variants of solutions for the manufacturer must have the same attributes, however, the values of these attributes may vary). This assumption is necessary due to the consolidation of the data contained in the structures, for example, has not occurred, the situation, in which one variant contains a date and the cost of delivery, and does not include the lot size, instead another variant consist of a date ant the lot size, and does not include cost of delivery. The choice of such „incomplete" variants is impossible (for example, you cannot select a supplier without knowing the cost of delivery).

The elaborated algorithm can be used to create the program module of consensus determining. Such a module can be placed in the structure of the SCM sys-

tem and started automatically after generated by system the different variants of solutions. After the execution of the algorithm, the solution that is the consensus is presented to the user, who shall decide on its implementation. The user is exempt from the need for continuous analysis of variants and making a selection from among the variants generated by the SCM system, time of determining destination variant is shortened, so making supply chain management becomes more flexible. In addition, the level of risk associated with the choice of a worst variant is reducing. As a result, a company may obtain a higher level of economic efficiency.

5. Conclusion

Supply chain management is related to the activities pursued by every organisation. In the past it was performed manually by people. Today, however, due to the nature of the modern economy, which is extremely turbulent, supply chain management without the use of IT systems is impossible. Support systems for such management, particularly SCM systems, permit the integration and coordination of product, information and cash flows between individual organisations being a part of a supply chain, which obviously affect the capacity of enterprises to adjust to the market demand properly. Employing SCM systems in supporting supply chain management also enables day-to-day market simulations, optimisation of delivery organisation, or definition of supply network limitations, which facilitates a prompt response to the emerging new demand on the part of customers. However, such systems operate effectively only on condition that responses to market changes are dynamic. The diversity of the criteria or methods of supply chain analysis adopted in SCM systems frequently lead to the situation where the system generates different solution variants including both incorrect and correct variants, which, for example, bear a very high risk level. In other words, there is a conflict of knowledge in this system. Choosing the best variant by the user is extremely difficult since it requires a detailed analysis of all variants, which naturally takes some time and, as a consequence, significantly reduces dynamism, thus decreasing the effectiveness of supply chain management. It is also uncertain whether the user will select the right variant even after conducting an analysis. It is often the case that – due to time limits related to the continuity of the supply chain management process – analysis is impossible and if the user relies on their experience only, the risk of choosing a wrong variant is high. Using consensus methods in resolving the conflict of knowledge, that is determining a variant to be then presented to the user, one based on the variants proposed by the system, may lead to shortening the variant determination time and to reducing the risk of selecting the worst variant. As a consequence, supply chain management might become more dynamic, which obviously influences the effectiveness of the operation of particular organisations and the entire supply chain.

Further research works should concern, among others, developing a program module determining a consensus in the SCM system and its review, developing consensus determination algorithms accounting for the differences in the structure of variant construction in individual chain areas such that, for example, it is possible to present different attributes of the variant in the case of the supplier, different in the case of the producer, and yet different in the case of the retailer. Consensus algorithms taking into consideration functional correlations between the structures of variants also need to be developed.

REFERENCES

[1] Allen, J.F. (1983) Maintaining *knowledge about temporal intervals, Commun.ACM*, no. 26, pp. 832-843.

[2] Barthlemy J.P. (1992) D*ictatorial consensus function on n-trees*, Mathematical Social Science, nr 25 , pp. 59-64.

[3] Christopher M. (2005) *Logistics and supply chain management, creating value-adding networks*, Third Edition, Financial Times Press, Paerson Education Limited.

[4] Dyk P., Lenar M. (2006) *Applying negotiation methods to resolve conflicts in multi-agent environments*, Zgrzywa A. (ed.), Multimedia and Network Information systems, MISSI 2006, Oficyna Wydawnicza PWr, Wrocław, pp. 259-268.

[5] Dyreson C.E., Soo M., Snodgrass R.T., (1995) *The data model for time*, Snodgrass R.T. (ed.), The SQL Temporal Query Language, Kluwer Academic Publish, Hingham M.A.,p.p. 327-346.

[6] Hernes M., Nguyen N.T. (2007) *Deriving Consensus for Hierarchical Incomplete* Ordered *Partitions and Coverings*, Journal of Universal Computer Science 13(2)/2007, pp. 317-328.

[7] Hernes M., Nguyen N.T. (2004) *Deriving Consensus for Incomplete Ordered Partitions*, Nguyen N.T. (ed.), Intelligent Technologies for Inconsistent Knowledge Processing, Advanced Knowledge International, Australia , pp. 39-56.

[8] Lu D. (2011) *Fundamentals of supply chain management*, Dr. Dawei Lu & Ventus Publishing ApS, bookboon.com.

[9] Nguyen N.T. (2006) *Using Consensus Methodology in Processing Inconsistency of Knowledge*, Last M. et al. (Ed.), Advances in Web Intelligence and Data Mining, series Studies in Computational Intelligence, Springer-Verlag, pp. 161-170.

[10] Podobnik V., Petric A., Jezic G (2008), *An Agent-Based Solution for Dynamic* Supply *Chain Management*, Journal of Universal Computer Science, vol. 14, no. 7, pp. 1080-1104.

[11] *Rutkowski* K. (2010) *Best Practices in Logistics and Supply Chain Management. The Case of Central and Eastern Europe*, Waters D. (ed.), Global Logistics and Distribution Planning, Kogan Page, London, pp. 14-30.

[12] Sitek P., Wikarek J. (2012) *Cost optimization of supply chain with multimodal transport*, Proceedings of the Federated Conference on Computer Science and Information Systems, http://fedcsis.org/proceedings/fedcsis2012/pliks/182.pdf., pp. 1111-1118.

[13] Siurdyban A., Møller C. (2012) *Towards Intelligent Supply Chains: A Unified Framework for Business Process Design*, International Journal of Information Systems and Supply Chain Management, No 5/2013, IGI Global, IGI Global, Hershey, New York, pp. 1-19.

[14] Sobieska-Karpińska J., Hernes M. (2010) *Value of information i distributed decision support system*, Pańkowska M. (ed.), Infonomics for Distributed Business and Decision-Making Environments: Creating Information System Ecology. IGI Global, Hershey, New York, pp. 153-176.

[15] Sobieska-Karpińska J., Hernes M. (2012) *Consensus determining algorithm in multi-agent decision support system with taking into consideration improving agent's knowledge*, Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS), pp. 1035-1040.

[16] Sobieska-Karpińska J., Hernes M. (2012) *Using consensus methods in knowledge conflicts resolving in supply chain management support systems*, Information Systems in Management, vol 1, no. 2, pp. 160-167.

[17] Sobieska-Karpińska J., Hernes M. (2013) *Distance functions between structure of variants in consensus determining process in supply chain management systems*, in: Dziechciarz J. (ed.), Econometrics , Wrocław University of Economics Press, Wrocław (in Press).

INFORMATION SYSTEMS IN MANAGEMENT

Information Systems in Management (2014) Vol. 3 (1) 40−52

COMPLEX COLOUR DETECTION METHODS USED IN SKIN DETECTION SYSTEMS

PIOTR MILCZARSKI, ZOFIA STAWSKA

Department of Theoretical Physics and Computer Sciences, University of Lodz, Poland

In this paper we present high computational complexity algorithms for detecting skin and non-skin colour. Because of their complexity they are not suitable for nowadays mobile devices but can be used in systems working on more demanding machines. The selection and implementation of algorithms gives accuracy about 80-90%.

Keywords: face identification, authentication, biometrics, detection of skin colour

1. Introduction

Face detection is used in a wide variety of tasks related to the recognition of people, including authentication and identification systems. The approach of using facial biometric systems seems very natural due to the ease and non-invasive acquisition of sample face. Detection of skin colour is usually the first step of automatic face recognition [1, 2, 15]. To date there is a wide range of skin colour detection methods. From the application point of view these methods in identification systems can operate on mobile devices. But the computational complexity is a very important parameter.

Searching for skin colours can be used in different areas like face and gaze tracking [7, 8]. The skin has also its optics and there are some anthropological differences [5, 6].

In the previous paper, *Algorithms and methods used in skin and face detection suitable for mobile applications* [31] we presented algorithms and methods of

image processing that can be used in image processing in mobile applications. Mobile applications can work on client side and can be used in mobile payment transactions i.e. using NFC.

In this section we would like to look at a more complex colour detection methods that can be used in systems identify the owner or user of the mobile device, where a portion of identification will be implemented at the application server side.

2. Parametric and nonparametric methods

The first of the algorithms that we want to analyze in the current paper is the one that identifies the colour in the images basing on standardization and classification of the RGB colour system and Bayesian classification, defined as:

$$P(skin|c) = \frac{P(c|skin)P(skin)}{P(c|skin)P(skin) + P(c|-skin)P(-skin)}$$

where *P (c | skin)* and *P (c |-skin*) skin and non-skin colours are taken from the histogram calculated by the following formula and normalization:

$$P_{skin}(c) = \frac{skin[c]}{Norm}$$

It gives a baseline for identifying skin. In order to further study the examples were used to train the histogram [11, 21, 22] and which has been illustrated by the following formula:

$$\frac{P(skin|c)}{P(-skin|c)} = \frac{P(c|skin)P(skin)}{P(c|-skin)P(-skin)}$$

where there was sought the difference between skin colour and the non-skin colour and these differences are described in the next formulas:

$$\frac{P(c|skin)}{P(c|-skin)} > \Theta, \quad \Theta = K \times \frac{1 - P(skin)}{P(skin)}$$

There was created a map on the basis of [16] that was designed to recognize only the skin and the non-skin using a database that you have been taught by

manually tagged pictures of their recognition. The created database has been tested in [21] and other systems qualifications pixels. However, they were not obtained expected results.

Distribution of colours can be represented by modelling the joint density of probability functions:

$$P(c|skin) = \frac{1}{2\pi|\Sigma_s|^{1/2}} \cdot e^{-\frac{1}{2}(c-\mu_s)^T \Sigma_s^{-1}(c-\mu_s)}$$

where c means a colour vector, μ_s and Σ_s parameters and matrices. Subsequently there can be obtained from the following formulas:

$$\mu_s = \frac{1}{n}\sum_{j=1}^{n} c_j \qquad \Sigma_s = \frac{1}{n-1}\sum_{j=1}^{n}(c_j - \mu_s)(c_j - \mu_s)^T$$

where n is the total number of examples of colour and the probabilities of the skin is "measured" in c [28]. An alternative to the above calculation was the distance c from μ_s that created the matrix Σ_s [9]:

$$\lambda_s(c) = (c - \mu_s)^T \Sigma_s^{-1}(c - \mu_s)$$

Gaussian method was used in [13, 15, 27, 29]. Gaussian method based on identification of the skin in the case of the following formula p meant skin; the remaining values are the values p_i and standardization.

$$P(c|skin) = \sum_{i=1}^{k} \pi_i \cdot p_i(c|skin)$$

Distribution of the skin with specific histogram and Gaussian model turned out to be insufficient [26]. Elliptical boundaries were proposed in that model. It gave much better results in the database [21]. This was determined by the formula:

$$\Phi(c) = (c - \phi)^T \Lambda^{-1}(c - \phi)$$

where colour samples with low frequencies and minimal data are removed.

Then the coefficient ϕ is estimated by the formulas:

$$\phi = \frac{1}{n}\sum_{i=1}^{n} c_i \quad \Lambda = \frac{1}{N}\sum_{i=1}^{n} f_i \cdot (c_i - \mu)(c_i - \mu)^T$$

$$\mu = \frac{1}{N}\sum_{i=1}^{n} f_i c_i \quad N = \sum_{i=1}^{n} f_i$$

in which *n* is an integer specific to the colour of the skin. Vectors *c* training pixels and f_i is the number of samples of skin colours vector *c*.

An important advantage of this method is to determine the skin by simple rules of eligibility. But there is a problem how to determine the colour of the skin appropriately. There was used a learning machine, which has overcome mentioned problems [30]. However, time devoted to both machine learning and recognition of image elements is significant.

3. Method based on distance map of colours DM [17]

In this method there was declared SSC - the standard colour of the skin by the vector. Its length - *n* determines the colour and the C vector (C_1, C_2 ..... C_n). Defined colour length CD as Euclidean distance between colours and SSC. Mathematically represented by the formula:

$$\sqrt{\sum_{i=1}^{n}\left(C_i - C_{i_s}\right)^2} \,/\, Sp_{(C_1,C_2,\ldots,C_n)}.$$

For example, RGB it will be counted by the formula:

$$\sqrt{(R - R_s)^2 + (G - G_s)^2 + (B - B_s)^2} \,/\, Sp_{(R,G,B)}$$

and the last assumption that the DM is presented in gray scale determined by the pixel CD and convert to gray scale linearly recognized colours:

$$DM(x, y) = \frac{d(x, y) - \min_{\forall x,y}(d(x, y))}{\max_{\forall x,y}(d(x, y)) - \min_{\forall x,y}(d(x, y))} \times 255$$

where *d(x, y)* is the pixel distance in the system (*x, y*).

Through the above deductions generated RGB frame for your skin in daylight

$$R > 95, \quad G > 40, \quad B > 20$$
$$Max\{R, G, B\} - Min\{R, G, B\} > 15$$
$$|R - G| > 15, \quad R > G, \quad R > B$$

and in the sideway light

$$R > 220, \quad G > 210, \quad B > 170$$
$$|R - G| \leq 15, \quad B < R, \quad B > G$$

If both conditions are true gray scale is formed and then combined into a single DM and calculated using the following formula

$$M(x, y) = Min\{Map_1(x, y), Map_2(x, y)\}$$

where x and y represent the coordinates of pixels in the image. The following algorithm illustrates the action of successive steps how to obtain

```
Algorithm Find_SSC()[17]

Input Parameter
H: The histogram of DM
ε: A threshold specifying satisfactory value of μ

Output parameter
C: The refined SSC for this image

Procedure
Set μ = first significant local maximum in H
Set Th = first significant local minimum in H, where Th > μ.
If components of H in [μ, Th] is close to right half of Gaussian
then
If μ < ε then
       Return C.
       Else
       Set C = Median of colour of pixel whose CD is μ
       Generate a new DM, M, with respect to C as SSC
       Generate H of M
       Go to step 1.
End If
Else
H does not represent any skin region
End if
```

Below there is the algorithm searching in the image seeds of the skin and non-skin pixels.

```
Algorithm Find_Seed()

Input Parameters
M: The refined DM of the test image.
T_L: A low threshold.
T_H: A high threshold.

Procedure
If M(i,j) ≤ T_L then
The pixel at position (i,j) is a skin pixel.
Else If M(i,j) ≥ T_H then
The pixel at position (i,j) is non-skin pixel.
Else
The pixel at position (i,j) is an undefined pixel.
End If
```

At last the algorithm of growing the area of skins [18, 19, 20] is

```
Algorithm Region_Growing

Input Parameters
G: Gradient magnitude of DM
S: skin seed points

Output Parameters
Seg: Segmented image of the size same as G

      1.    Label each pixel in Seg as “skin”, “non-skin” or
      “undefined” according to S and NS.
      2.    Add the neighbouring pixels of labelled region in the
      respective queues according to their gradient magnitude
      levels.
      3.    While all queues are not empty do
            a.    Pick a pixel p from the first available nonempty
            queue of the ordered queue according to priority of
            queues
            b.    If p has similarly labelled neighbours, then it
            is labelled as them. Otherwise, it is labelled as a
            boundary pixel.
            c.    For each undefined neighbouring pixel q of p
                 i.      If q is not already added in queue, add q
                   in the respective queue according to its
                   gradient magnitude level.
```

The method achieves excellent results and is able to find a light skin colour on the picture, regardless of the origin of the person shown on the picture, the

background, lighting, and environments. It is effective for identifying areas without skin. In assessing effectiveness used, there was used Compaq database of the skin and non-skin. It includes an appropriate number of images of the skin and slow. 4000 randomly selected images containing areas of the skin and 5500 images slow to assess existing solutions. In addition to testing a randomly selected set of sequential images 500 consisting of 62,100,260 pixels, 9,859,733 pixels in the skin and 52,240,527 pixels slow. All the data obtained in Table 3 are the average values based on simulation 500 test images.

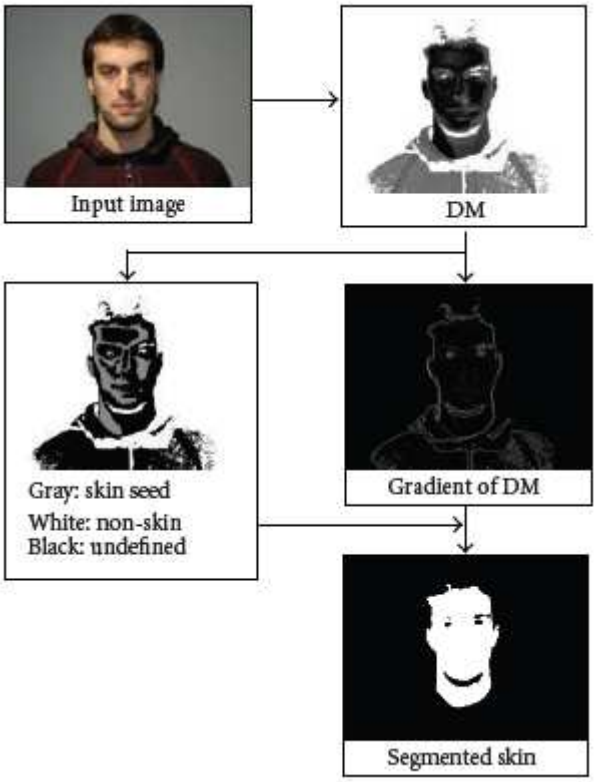

Figure 1. Algorithm workflow [17]

Table 1. Effectiveness of recognition methods of skin and non-skin colours [17]

Method / classifier	CDR (%)	FDR (%)	CR (%)
Traditional RGB method	81.2683	23.7099	77.0412
DM – described method	**89.9749**	**9.2695**	**90.6165**
Bayesian classifier [4, 14]	83.9234	10.9183	88.3034
Multilayer perceptron classifier [12]	83.3306	11.5401	87.6861
Colour consist with implementing neural networks [13]	85.6037	10.6809	88.7585
Segment- and edge-based refinements of Bayesian classifier [14]	82.6245	10.4442	88.4416
Principal feature analysis, PFA and Markov random field , MRF based methods [15]	83.9304	10.8703	88.3453

4. Pixel classification method [3]

Another algorithm is developed at the base of face detector [23] and the method of classification pixels [24]. With both elements it creates a map of the skin, not only for the type of Europeans but also for other anthropological types to form segments by the pixels and collects them in the LCH colour. This can be used in variety of methods i.e [10]. Using the method each pixel classification [24] is defined as the probability function of the skin:

$$P(skin|l,c,h) = Z\, e^{-\left(\frac{(l-\mu_l)^2}{2\sigma_l^2}+\frac{(c-\mu_c)^2}{2\sigma_c^2}+\frac{(h-\mu_h)^2}{2\sigma_h^2}\right)}$$

where l, c, h are the coordinates of the pixel map LCH, $\mu_l = 181$, $\mu_c = 32$, $\mu_h = 34$, $\sigma_l = 30$, $\sigma_c = 11$, $\sigma_h = 8$, and Z is a normalization factor. Probability of detection of the skin is compatible with a high factor, small, if it is low. The following are two methods for detecting the skin, wherein the first presupposes that the pixel is a skin:

$$w_x \frac{(x-\mu_x)^2}{2\sigma_x^2}, w_y \frac{(y-\mu_y)^2}{2\sigma_{yx}^2}, w_l \frac{(l-\mu_l)^2}{2\sigma_l^2}, w_c \frac{(c-\mu_c)^2}{2\sigma_c^2} \quad and \quad w_h \frac{(h-\mu_{xh})^2}{2\sigma_h^2}$$

Based on the weighted average over 1000 adjacent pixels in space of five functions. Where l, c, h means the position of the pixel in the x, y.

Another method assumes a Gaussian probability model in the space of five functions:

$$P(skin|l,c,h,x,y) = Z \cdot e^{-\left(w_l\frac{(l-\mu_l)^2}{2\sigma_l^2}+w_c\frac{(c-\mu_c)^2}{2\sigma_c^2}+w_h\frac{(h-\mu_h)^2}{2\sigma_h^2}\right)} \cdot e^{-\left(w_x\frac{(x-\mu_x)^2}{2\sigma_x^2}+w_y\frac{(y-\mu_y)^2}{2\sigma_y^2}\right)}$$

where Z is the normalization factor and the relative weight of each function. Then, to create a map of the skin, use the following function, which will give us a value indicative of the skin in the study area.

$$P(skin|c,h) = Z \cdot e^{-\left(w_c\frac{(c-\mu_c)^2}{2\sigma_c^2}+w_h\frac{(h-\mu_h)^2}{2\sigma_h^2}\right)}$$

The framework of the adaptation maps the colour is specified by the above formulas. They allow automatic segmentation of the skin pixels and then creating their prototype. Colour correction is performed using LAB colour areas, which represent the Cartesian cylindrical LCH

$$a = c \cdot \cos(h)$$
$$b = c \cdot \sin(h)$$

where *a* and *b* will be the prototype of skin in the plane of exposure CIECAM - UCS [25], *image a* and *image b* are the coordinates of computer memory and a_{orig} and b_{orig} are the coordinates of pixels. The variable *k* is the colour correction factor typically in the range of one.

$$a^*_{new} = a^*_{orig} + \Delta I_a kP\left(a^*_{orig}, b^*_{orig}\right)^\gamma$$
$$b^*_{new} = b^*_{orig} + \Delta I_b kP\left(a^*_{orig}, b^*_{orig}\right)^\gamma$$

At this level, the probability of a pixel is calculated skin. This algorithm, using information from the face detection and skin colour model and machine learning - learning model for every person of colour in the image and uses these models to calculate the maps of the skin. Each of the tested pixel is assigned probability of being skin.

The test consisted of two parts using outdoor images [3]. The first one was for training and algorithm tuning and contained 200 images from the Berkeley segmentation dataset [18], and was tagged by the authors [3]. The second one was used for testing and contained 196 images, and that was tagged by an external person. The results of the test are as follows:

- 80% of the skin pixels were tagged correctly,
- 21% was wrong detection.

5. Conclusion

In the previous paper, *Algorithms and methods used in skin and face detection suitable for mobile applications* [31] we presented image processing algorithms and methods of that can be used in authentication systems working on mobile client side. Their accuracy was about 80%. Their advantage was the low computational complexity and that is why they can be used in mobile applications.

Presented in this article, the skin colour detection methods generally provide better results in comparison with the methods being considered in the first part of the article. Their accuracy reaches 90% but their computational complexity is much higher than described in the paper mentioned above. These are methods can perform better in more complex situations where there are some disturbances that adversely affect the process of identifying the pixels representing the human skin.

In the case of skin colour detection algorithms in identification / authentication systems accuracy is about 80% and it may be insufficient. Therefore, there is a great demand for algorithms that provide greater accuracy of results. In this article, we presented a few selected methods that meet this criterion. These algorithms detect skin pixels in the image with the accuracy of about 90%. Such a result is generally satisfactory from the standpoint of further processing of subsequent image for the purpose of identification / authentication. Unfortunately, increasing the accuracy of these methods is associated with a fairly significant increase in computational complexity. These methods carry out a series of relatively complex mathematical transformations. Thanks these transformations we can extract additional information from the image.

The increase in computational complexity is so significant that it prevents us from using these solutions in applications running on the client side in currently available mobile systems/ devices. However, this does not mean, that they are completely unusable in our applications. Many identification / authentication systems allow us to perform some of the necessary calculations in this process on the server side. The server does not require such restrictive limitations on the computational complexity as for the algorithms running on mobile devices. Usually, the system has enough computing power to perform even the most complex calculations in real time. It seems that such a solution would also provide a very high quality of recognition and satisfactory operation time of the application. This assumption is based on the fact that the network connection has got sufficient bandwidth.

In addition, we expect that the computing power of mobile devices will continue to increase. Perhaps in the near future average commercially available mobile device will be able to cope with the complexity of the algorithms similar to those presented in the article.

In the case of authentication or identification systems the reliability of the result is so important that we should strive to use algorithms that give the highest accuracy, even at the expense of the significant increase in computational complexity.

From the presented tables it can be derived that the most promising results can be obtained from the methods using the distance map, DM. Therefore, it will be one of the methods used by us in further research on facial recognition systems operating on mobile devices using server-side processing.

REFERENCES

[1] Crowley J. L., Coutaz J. (1997) *Vision for Man Machine Interaction*, Robotics and Autonomous Systems, Vol. 19, 347-358.

[2] Kjeldsen R., Kender. J. (1996) *Finding Skin in Color Images*, Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 312-317.

[3] Nachlieli H., Bergman R., Greig D., Staelin C., Oicherman B., Ruckenstein G., Shaked D. (2009) *Skin-Sensitive Automatic Color Correction*, SIGGRAPH, New Orleans.

[4] Qian Chen, Haiyuan Wu, Masahiko Yachida (1995) *Face detection by fuzzy pattern matching*, In Proc. of Fifth Intl.Conf. on Computer Vision, 591–596, Cambridge.

[5] Symon D'Oyly Cotton, Ela Claridge (1996) *Do all human skin colors lie on a defined surface within LMS space?* Technical Report CSR-96-01, School of Computer Science, Univ. of Birmingham, UK.

[6] Van Gemert M. J. C., Jacques S. L., Sterenborg H. J. C. M., Star W. M. (1989) *Skin optic*,. IEEE Trans. On Biomedical Engineering, 36 (12): 1146–1154.

[7] Jebara T. S., Pentland A. (1997) *Parameterized structure from motion for 3d adaptive feedback tracking of faces*, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 144–150, San Juan, Puerto Rico.

[8] Schiele B., Waibel A. (1995) *Gaze tracking based on face-color*, In Proceedings of the International Workshop onAutomatic Face- and Gesture-Recognition, 344–349, Zurich, Switzerland.

[9] Terrillon J-C., David M., Akamatsu S. (1998) *Automatic Detection of HumanFaces in Natural Scene Images by Use of a Skin Color Model and of Invariant Moments.* In Proc. of the Third International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 112-117.

[10] Zeng H. (2011) *Preferred skin colour reproduction.* PhD thesis, University of Leeds.

[11] Zarit B. D., Super B. J., Quek F. K. H. (1999) *Comparison of five color models in skin pixel classification*, In ICCV'99 Int'l Workshopon recognition, analysis and tracking of faces and gestures in Real-Time systems, 58–63.

[12] Poynton C. A. (1995) *Frequently asked questions about colour.* In: *ftp://www.inforamp.net/pub/users/poynton/doc/colour/ColorFAQ.ps.gz*

[13] Ahlberg J. (1999) *A system for face localization and facial feature extraction*, Tech. Rep. LiTH-ISY-R-2172, Linkoping University.

[14] Phung S. L., Bouzerdoum A., Chai D. (2002) *A novel skin color model in ycbcr color space and its application to human face detection*, In IEEE International Conference on Image Processing (ICIP'2002), Vol. 1, 289–292.

[15] Hsu R.-L., Abdel-Mottaleb M., Jain A. K. (2002) *Face detection in color images*, IEEE Trans. Pattern Analysis and Machine Intelligence 24, 5, 696–706.

[16] Brown D., Craw I., Lewthwaite J. (2001) *A som based approach to skin detection with application in real time systems*, In Proc. of the British Machine Vision Conference.

[17] Abdullah-Al-Wadud M., Chae O. (2007) *Region-of-interest selection for skin detection based applications*, In Proceedings of the International Conference on Convergence Information Technology (ICCIT '07), pp. 1999–2004, Gyeongju, Korea.

[18] Bera A. K., Jarque C. M. (1980) *Efficient tests for normality, homoscedasticity and serial independence of regression residuals*, Economics Letters, vol. 6 (3), 255–259.

[19] Bera A. K., Jarque C. M. (1981) *Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo Evidence*, Economics Letters, vol. 7, no. 4, 313–318.

[20] Judge G. G., Hill R. C., Griffiths W. E., Lutkepohl H., Lee T. C. (1988) *Introduction to he Theory and Practice of Econometrics*, JohnWiley & Sons, New York, NY, USA, 3rd edition.

[21] Jones M. J., Rehg J. M. (1999) *Statistical color models with application to skin detection*, In Proc. of the CVPR '99, vol. 1, 274–280.

[22] Chai D., Bouzerdoum A. (2000) *A bayesian approach to skin color classification in ycbcr color space*, In Proceedings IEEE Region Ten Conference (TENCON'2000), vol. 2, 421–424.

[23] Viola P., Jones M. (2004) *Robust real-time face detection*, International Journal of Computer Vision, vol. 57, no. 2, pp. 137–154.

[24] Martinez B. O. (2004) *Semantic retrieval of memory color content*, Ph.D. dissertation, Universitat Autonoma de Barcelona.

[25] Luo Ronnier G. C., Li C. (2006) *Uniform colour spaces based on ciecam02 colour appearance model*, Color Research and Application, vol. 31, no. 4, 320–330.

[26] Lee J. Y., Yoo S. I. (2002) *An elliptical boundary model for skin color detection*, In Proc. of the 2002 International Conference on Imaging Science, Systems and Technology.

[27] Yang M.-H., Ahuja N. (1998) *Detecting human faces in color images*, In International Conference on Image Processing (ICIP), vol. 1, 127–130.

[28] Menser B., Wien M. (2000) *Segmentation and tracking of facial regions in color image sequences*, In Proc. SPIE Visual Communications and Image Processing 2000, 731–740.

[29] Saber E., Tekalp A. (1998) *Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions*, Pattern Recognition Letters 19, 8, 669–680.

[30] Gomez G., Morales E. (2002) *Automatic feature construction and a simple rule induction algorithm for skin detection*, In Proc. of the ICML Workshop on Machine Learning in Computer Vision, 31–38.

[31] Stawska Z., Milczarski P. *Algorithms and methods used in skin and face detection suitable for mobile applications*, in preprint.

INFORMATION
SYSTEMS IN
MANAGEMENT

Information Systems in Management (2014) Vol. 3 (1) 53–63

MANAGEMENT INFORMATION SYSTEMS IN POLISH RAILWAY TRANSPORT IN THE CONTEXT OF THE TECHNOLOGICAL AND THE ORGANIZATIONAL CHANGES

MICHAŁ HENRYK RUDOWSKI

Institute of Computer Science, Warsaw University of Technology
PKP Informatyka Spółka z o.o.

Paper presents the implementation problems of information management systems in the railway industry with an analysis of the current situation, the needs of companies and benefits of the implementation. It takes into account the area of financial management, a range similar to the other companies, as well as the area of operational management specific for their activities. These issues are presented in the context of technological changes taking place since the beginning of the XXI century in the information management systems and organizational changes in the PKP Group with the transport operations and railway infrastructure management separation, consolidation of the PKP Group and the changes connected with the preparation for the privatization of some of the group entities.

Keywords: information management systems, railway industry, PKP Group, information technology

1. Introduction

This paper presents the historical, technological and industrial conditions of management systems in the Polish State Railways. Even during the communist regime (which was not the best time in the management of business organizations) in the absence of now widely available and necessary tools it was possible to suc-

cessfully implement systems supporting management processes. Thay allowed not optimal, but effective management of a giant enterprise, which were Polish State Railways. Selected whole railway network systems implemented in the first decade of transition before the embargo on information technology has been canceled are presented. Also systems implemented and deployed after 2001 in PKP Group companies are shortly described. Moreover the changes in the PKP Group and challenges in the area of management systems are pointed out. The factors that allows effective and long-term operations of the management systems, on the example of material management system GOSMAT are indicated in the summary.

2. Historical, technological and industry conditions of the management systems implementation in the polish railways

Railways as a great and complicated organization operating throughout the country always had high demands, therefore management systems became very early on. IT Department of the Polish State Railways (PKP) comes from the Central Statistics Office established on 1 January 1958. It implemented statistical calculations using the tabulating machines. The results were used in the assessment of the economic situation and in making management decision. In 1962 Central Statistics Office has been transformed into the Central Office of Mechanization and Automation of Statistical Calculations (COZO) which was directly under the Ministry of Communication. At the same time Bull series 300 MCT machines were introduced into operation. Those machines formed a connection between analytical machines and electronic computers. Computers appeared in the 1960s. in the form of ICL series 1900 machines and then ODRA series 1300 machines manufactured under license by the Wroclaw factory ELWRO. These computers in subsequent versions PKP bought at the beginning of the 1980s. Some of them had been used until the beginning of the 21st century. In 1976 COZO (with subordinated to it centers in Warsaw, Łódź, Sosnowiec, and Olsztyn) was transformed into Central Information Technology Office of Railways (COIK). Then the Management System Design and the Control System Design Services have been created in the COIK and processes of modern business management systems development and operational management of rail traffic systems have been launched. In addition to the COIK organizational structures between 1974 and 1979 in the Regional Directorates of Polish State Railways had been created Regional IT Centers [1].

When the 3rd generation computers had been implemented the whole railway network systems corresponding to the most important challenges of enterprise management have been developing and implementing. Until 1980s PKP management systems like in the majority of other companies were autonomic and scattered as in the general management area called for the purposes of this article *back office*

area as well as in the operational management area called *front office area*. Dispersal and diversification stemmed from technical and organizational reasons. Technical reasons were the lack of computer networks. Only local terminals and telexes with limited bandwidth were available. In addition to the central structure of the COIK Polish State Railways company had its own Regional IT Centers, which in a number of ways were responding to the demands in the management systems area. Their tasks had not been properly coordinated. The effect was that they developed different systems solving similar problems but experiences in solving the problems could not be used by others. Different business units had used different technical solutions and tools. This situation lasted until 1992, when the Information Technology Department of Polish State Railways was created and Regional IT Centers started working in that structure. Main tasks of the Information Technology Department were: designing, implementing and operating of traffic management system and financial management system (SKPZ – system), organization and coordination of the other IT applications, agreeing IT applications projects of organizational units, coordination of statistics issues, the methodology and standards development for the information systems [1].

Back office systems should not be different from those used in other businesses. The specificity of the railway company should not affect the requirements of systems like: HR, payroll, financial and accounting, material management, car fleet management and so on. The reality was quite different. Historical circumstances, specifics of the company, autonomy, strong position of trade unions among many other things impacted management systems characteristics, so that they will not only meet not only the general requirements, but also numerous railway industry ones. The best example are the Engine Houses which had to take into account the hundreds of items constituting the train driver salary. More reasonable were the diversities of systems in the "front office" area that is systems related to the characteristics of the company. Back then Polish State Railways were the largest corporation in terms of employment, area of operation, ownership of property and the diversification of activities. Their activities were including not only the activities from the scope of the maintenance and repair of the railway infrastructure, manage the infrastructure and carrying out the movement of domestic and international passenger and freight trains of all classes and types but also support in the energy, telecommunications, information technology, facilities repairs, repair, health, security, design units, as well as conducting scientific research. The additional complications were technological restraints. By the end of the 1980s there were no country-wide computer networks but the company had working throughout the country and train movement processes covered the whole territory of Poland. Systems which were developed in these conditions used telephone or telex network. In some cases as a method of data transmission was used collecting entered data and transferring them to the Data Center on different available in a given

period media or just printing them and transferring to the destination in the printed form. With all these limitations multi-year operations of country-wide management information systems were a great success.

3. Selected management support systems implemented in the 1980s

Many management support systems had been developed until the end of the 1980s. Few of them are interesting because of particular importance for the company and the fact they were leading solutions in Europe in this area. Those systems were CETAR, BEWAG and INTERWAG in the front office area and GOSMAT in the back office area [1].

CETAR is the **c**entral system for calculating and settling debts for freight. Its concept was established in the 1970s and the main features include: computing dues, automatic billing of PKP clients, settling debts with foreign railway authorities, the making-out of physical statistics, financial statistics and analysis to illustrate the cargo freight. In a sense, it had become a victim of its quality, because the client was not interested in its modernization. Only the current adaptations to legislative changes and organizational changes in the PKP enterprise and in the company PKP Cargo were done. The System ceased to existence in 2007. With it the last machine of ODRA/ICL type which was supporting management systems was turned off.

BEWAG is the central records of wagons. Its goal was to operate records of railway wagons. For each wagon were recorded: operating and technical characteristics, financial and accounting information, repair data and archived data. The database established on the basis of wagon status defining cards everyday was updated by telegraph messages about shopping, deletions or by monthly reports about included and excluded wagons according with agreements with their owners. The results were submitted to the terminal screens, by telegram or in the form of prints documenting income, payments, depreciation and containing a statement of quantity-value wagons required for the approval of the balance sheet in area of PKP assets. Draft plans for wagon repairs were also generated.

INTERWAG. The aim of the system was to provide financial and statistical information about the exchange of wagons in international traffic. The information concerned the settlement rents and included comparative lists to check the correctness of the calculation of the rents and the data for the analysis of wagons circulations across the borders of the state. Input data were derived from lists established at border stations after train acceptance or transfer or passed by radiotelephone inventory reports. The system gathered over 10,000 messages a day. Users received billing statements daily, every ten days, monthly, quarterly and on annual basis. The statements were delivered to the Central Office of Foreign Accounts (CBRZ) and they included: wagon aid accounts, settlement accounts for foreign railways,

sealed PKP wagons turnover, accounts for the use of wide track wagons, reports of the time stay of wide track wagons in handling areas, exchange load reports from border stations, list of foreign wagons on PKP network, number, status and stay time of PKP wagons abroad, exchange of wagons reports according to border stations.

GOSMAT is the central material management system, which dates back to the 1960s, when it confined to the records of status and material flows. Since 1976, the system has been implemented on the entire rail network. The aim of the scheme is to provide users with the necessary accounting information and financial information of the material circulation characteristics in both product structures and organizational structures on the selected level of aggregation and management information needed for decision making. After multiple modernizations system is still operating and it supports a wide range of Group companies and other entities. GOSMAT currently uses Oracle Database and SAP Business Objects. A few years ago it was considered as one of the best material management systems with the title "Market leader in 2006, Euro Leader" awarded by the magazine *New Industry* (*Nowy Przemysł*) and it was highlighted at rail industry conferences.

4. Changes related to the transformation of the political system and access to new technologies

Since 1989, the political changes made new methods and technologies available because embargo on the modern information and communication technologies was lifted. In 1989, part of a World Bank loan granted to PKP was spent on IT projects. To enable the construction of modern enterprise management systems computer network is required. Packet network according to CCITT recommendation X.25 was selected to implement. In 1992 contract for the supply of equipment and the construction of the network KOLPAK was signed with the American corporation Sprint**.** KOLPAK was first countrywide packed network in Poland. It was used not only by the railway company, but it to other state institutions took advantage of it. The aim of the network was to enable the implementation of projects in the field of management called SKPZ (Operation and Management Information System of the Enterprise). SKPZ consisted of two main components **OMIS** (Operational Management Information System) and **FMIS** (Financial Management Information System) that with the KURS-90 system supporting activities in the field of passenger transport were supposed to be a new quality. OMIS and FMIS were integrated and centrally managed systems using state-of-the-art technology at that time. The goal of OMIS project was to create operation management system by PKP IT Department on our own. Applications of the system has been implemented and modernized since 1995. The aim of FMIS project was to purchase and adapt integrated financial management system which used the same

hardware and software platform like OMIS (operational system OpenVMS and Oracle Database). In the SKPZ project special attention was paid to ensure the operation and business continuity, hence the choice of hardware platform AXP Alpha systems in the OpenVMS cluster environment. Given that the KURS-90 system also worked with high reliable infrastructure Tandem and also X.25 network was characterized by high availability, in the 90s PKP has become a national leader in robust country-wide management information systems. Despite the technical reliability of the hardware and software platform created opportunities were not used due to the fact that the OMIS system design and adaptation to the PKP needs and the implementation of FMIS system lasted too long. System hardware platforms were no longer so important for software vendors (Oracle), and equipment was no longer sufficient for a large company. The OMIS project isolated applications related to the timetable (RJ) because performance of Alpha computers on 1994 proved to be insufficient for these applications. Thus, in 90's following key management systems replaced obsolete ones: KURS-90 in the ticketing and reservation, OMIS on the description of the railway rolling stock records, management and operation of commercial wagons freight, RJ in the creation and distribution of timetables, and FMIS in the range of company financial services [2].

All of the above-mentioned systems have used a X.25 network KOLPAK, and since 1999 the IP network implemented by IT Department of PKP with the DEC and Cisco routers and SDH 2 Mbps links. It was one of the first country-wide private IP network.

5. Selected systems implemented in PKP group companies

January 1, 2000 pursuant to the Act of 8 September 2000 on commercialization, restructuring and privatization of the state enterprise "Polish State Railways" [3] national railway company was converted into a joint stock company. In October 1, 2001 it was divided into a number of companies. This law was inspired by COUNCIL DIRECTIVE of 29 July 1991 on the development of the Community's railways (91/440/EEC) [4] requiring the separation of passenger and freight transport roles of the infrastructure management role. The following entities were created: PKP Polish Railway Lines - the infrastructure manager, PKP Cargo SA - the main freight carrier and logistics operator, PKP Intercity - qualified country-wide passenger carrier, PKP Przewozy Regionalne - a passenger carrier offered country-wide local, regional and inter-regional trains except qualified to operate a PKP Intercity company. Also brought to life a number of smaller entities: carriers operating passenger (WKD, SKM), and freight (LHS) which benefited from a dedicated infrastructure. In addition, several established companies with infrastructure: PKP Energetyka, Telekomunikacja Kolejowa and PKP Informatyka. PKP Informatyka took over the information technology services for the companies, but

without a monopoly position. Only two of the railway companies have implemented FMIS system: Telekomunikacja Kolejowa and PKP Informatyka. They used the previously conducted studies for FMIS system and in a short time made the implementation of Oracle Applications. The reason for the resignation from FMIS were not the technical or functional deficiencies that have been overcome. More important was the fear of assigned costs incurred in the implementation in the huge PKP enterprise to small companies. Company, which almost as quickly implemented **Oracle Applications** now known as **Oracle e-Business Suite** was the LHS. Oracle Applications were the first of ERP system implemented in the companies formed from the PKP. Another company that has implemented this system in 2007 is PKP Intercity. In subsequent years PKP Intercity has implemented data warehouse system Oracle Business Intelligence (BI) [5]. Oracle e-Business Suite in PKP Intercity has passed a serious test, when as a result of the reorganization of group companies PKP Intercity being exclusively engaged in the qualified passenger carriage took over from PKP Cargo locomotives to operate the passenger carriage, engine houses and repair facilities and from the PKP Przewozy Regionalne inter-regional passenger services. As a result of this reorganization, PKP Intercity increased several times in the area of the duties, assets and employment. The reorganization has not improved financial results of the company, but the system has passed this exam. Only the hardware infrastructure had to be upgraded because it was designed for a much smaller company. The proof that the system operated well for PKP Intercity, is running in 2012 project INTER12 - implementation of the new version of Oracle e-BS R12 and the new version of Oracle Business Intelligence systems.

The largest PKP Group companies have chosen SAP ERP system. The first implementation of SAP ERP in the companies of the group took place in PKP Regional Services. This was before it municipalization and leaving of the PKP group. Another implementation has been carried out in PKP Cargo, including a very difficult area for Polish railway HR. It has been really a big challenge, given the scale and complexity of the company payroll. The largest in the scope and scale, implementation was carried out in several stages at PKP Polish Railway Lines. Implementation has begun with the selected modules of approximately 300 users, and during next three years expanded the by implementing appropriate new modules and hardware components. Now the system supports more than 2,000 users. After the implementation of the first group of modules SAP ERP data warehouse SAP BW and SAP Business Objects reporting platform have been deployed. It is one of the largest SAP system implementations in Poland. In addition, the SAP ERP system has been implemented in the following companies: PKP SA (parent company of the group) and PKP Energetyka. Larger companies have decided to choose SAP system whereas smaller companies remained Oracle eBS system

and Oracle Business Intelligence. Oracle BI system is also being implemented in PKP Cargo.

Significant changes also occurred in the area of core activities management support. PKP PLK has introduced a SEPE system (Registration System of Operation). It maintains and displays the status of real time train traffic throughout the rail network managed by PKP PLK. It is developing SKRJ system (Timetable Constructing System), which will replace the RJ system. In 2007 implementation of the system SILK (Railways Information System) started. SILK is a typical GIS. SILK is carried out using Bentley and Oracle Spatial tools by PKP Informatyka with SHH - company specializing in GIS. Since 2008, the modules of: real estate, sections of railway lines with functionality of LRS (Line Reference System), internal railway maps and public railway maps has been implemented.

PKP Cargo has become the main beneficiary of the OMIS system whose applications support services freight and rolling stock. Applications of the OMIS system between 1995 and 2007 replaced earlier systems including BEWAG, INTERWAG and CETAR. Now the system consists of:

- Registration Application Group (ETP - traction registry, EWAG – wagons registry, GPW - private and rented wagons management)
- Freight Shipments Commercial Service Support System (OHPT - commercial freight shipments service, UMAK - acquisition agreements, ON - dues calculations, RZK - international settlements),
- Wagon Management and Freight Shipment Tracking Application Group (KPS – station operations guiding, ZPWO - foreign empty wagons returning, WIP - wagons and trains),
- Database of whole rail network description connected with POS application operated by PKP PLK.

Now PKP Cargo has started the implementation of Maintenance and Rolling Stock Management System based on IBM Maximo.

The second freight carrier PKP LHS has implemented INFO-LHS system – dedicated, integrated freight management system by PKP Informatyka.

For PKP Intercity basic sales system is KURS-90 which has been enriching from 2008 with capabilities of KURS-2008 system which is still being developed.

All freight and passengers carriers are users of the RJ or SKRJ systems, placing orders and running the trains in accordance with the timetable.

6. PKP group changes and challenges in area of management systems

Since 2012, a lot of changes occurred on the management of PKP Group companies. New managers put a lot of emphasis on intensifying joint operations of amongst group companies in order to reduce costs by exploiting economies of scale. Unification would bring many benefibeetween ERP systems. At the front

office area most companies have to operate their own system to suit the specific business requirements. On the other hand, the group will experience further changes. PKP PLK is leaving of PKP Group as the state infrastructure manager, PKP Cargo is partially privatized. Several other companies are awaiting privatization. Expecting significant changes companies invest only in the most urgent task.

All management systems require modernization and integration. Many systems in one organization existed side by side. Multiple interfaces were built between them but some of the needs are still unmet both quantitatively as well as qualitatively. There is a need for such integration of systems within both front office and back office areas work together as the client and the server. These are the known Web Services integration and data buses. Many systems include them or are adapted to cooperate with them (SAP, Oracle, IBM, KURS2008) but there is lack of appropriate standardization. The big tasks are the implementations of document management systems to implement business processes in such a way that they download data from and insert data to appropriate management systems. Most companies have already started work in this area.

In the front office area of PKP PLK the most important is development of rail network and real estate management system which could locate and list all investments with precise spatial location. Mmore accurate representation of the rail network is required (tracks instead of lines).

For freight carriers important issues are: cooperation with the clients and other carriers, the electronic exchange of information and the accurate information about the current location and the time of delivery of freight shipments. Implementation of the electronic consignment note system was completed in September 2013. The key issue now is to optimize the maintenance of rolling stock with IBM Maximo system.

In area of passenger services, the most important thing is to develop a common ticket sales platform where everyone could buy ticket regardless of the initial and destination stations, regardless of the number of direct trains and the number of carriers, at the lowest price, taking into account all available promotions and discounts, using any method of payment. PKP Informatyka declared to develop such platform.

At the end a few words about the technological challenges. Requirements, as to the availability and efficiency of processing have led to a situation that the systems are operating in many different sometimes quite exotic platforms. The development and dissemination of virtualization, high availability and performance maintenance techniques allows implementation of the management systems on the standard, uniform of hardware and software infrastructure. Using the features offered by consolidation, virtualization and automation it is possible to get huge cost savings. Consolidated and virtualized systems with automating maintenance tasks enable usage of cloud computing. The largest group companies can afford to im-

plement private cloud. It has many advantages but the operation of one cloud instead of multiple cloud solution will be much more economical. Smaller companies should use the services provider specialized in cloud computing services delivery. In author's opinion it is worth for the group that one company has been supplier of cloud computing services for all group companies and for other entities.

Some modern information systems, for example Oracle Database 12c are prepared to share the resources to the various players.

7. Summary

The article summarizes the history of information management systems in Polish railways. It is interesting what factors influenced that certain systems were replaced soon and some were exploited even for decades. Let us examine GOSMAT system, which has been changed technologically and architecturally and which has been used for more than 40 years. This is not an isolated situation, as systems of leading manufacturers (Oracle, SAP) also have a similar longevity. Let's look at the main features of this system.

Its purpose is to collect and share data on trading material for many independent business entities that use a single index of the material. It is implemented through a central database with quantity and value information. The basic features of the system include:

- Presentation of current and reliable data concerning the economy of materials at different management levels and the required level of aggregation;
- Quick and reliable access to information about current state of stocks in the company organizational units in the required form;
- Easy implementation of new reports at the request of the Customers;
- Provide user-defined reports within the Intranet or the Internet from anywhere in the world according with the requirements of security;
- The possibility of a complete audit of the users activities;
- Provide for the exchange of data with ERP systems;
- No need to install, configure and administer the workstation software;
- Any operating system and low requirements on the workstation.

The factors that determined the longevity and still high evaluation of system (awards and honors in recent years), according to the author are:

- A good definition of the requirements and correct implementation, taking into account the possibility of making functional and non-functional changes;
- Current and clear system documentation;
- Flexible parameterization of the system;
- The use of appropriate tools and platforms (Linux, Oracle Database, SAP Business Objects) and their appropriate use;

- Switching to new technologies (changing system platform, database, application, virtualization etc.);

- Quickly adapt the system to changing user needs;
- Maintaining a dedicated team responsible for the development and operation of the application.

Even compliance with these requirements does not guarantee continuity of operation. With the implementation of ERP platforms, customers deploy applications that are running as a part of the systems. According to the author, GOSMAT system can only survive as a part of the ERP system created by PKP Informatyka for the SMB sector companies. GOSMAT as a stand-alone system is likely to disappear from the offer within a few years.

The paper shows that while the management systems made lot of work, but because of the great organizational changes resulting from the Act [3] and subsequent reorganizations, changes of plans, changes of ownership, the separation of regional services to the local authorities current state of information management processes and the systems supporting them is not satisfactory in terms of functionality and technology. Even when systems provide excellent support in terms of availability and performance, they are not the best solution in economic terms. According to the author, information management systems are the subject that PKP Group companies should invest to enable faster decision making at lower cost through the use of new technologies and market opportunities.

REFERENCES

[1] Wielądek Adam F. (2008) *From abacus to… 50 years of IT in railways*, PKP Informatyka, Warsaw, Poland (in Polish).

[2] Chadrych Stanisław (2000) *Finance Management System FMIS as enterprise management tool*, Scientific and Technical Conference "Kolej na Informatykę", 14-15 December 2000, Jelenia Góra, Poland (in Polish), 49-60.

[3] *The Act of 8 September 2000 on commercialisation, restructuring and private-sation of the State Enterprise "Polish State Railways"*, Dziennik Ustaw 2000 nr 84 poz. 948 (in Polish).

[4] *COUNCIL DIRECTIVE of 29 July 1991 on the development of the Commu-nity's railways* (91/440/EEC), 1991, utk.gov.pl/download.php

[5] Krysztofiak Katarzyna (2007) *Oracle e-Business Suite as a friend of PKP Intercity*, Scientific and Technical Conference "PKP in the EU - lights and shadows", 19-21 September 2007, Jastrzębia Góra, Poland (in Polish), 51-61.

INFORMATION
SYSTEMS IN
MANAGEMENT

Information Systems in Management (2014) Vol. 3 (1) 64–76

SECURITY MANAGEMENT IN THE *H-H TYPE* OF THE ORGANIZATIONAL STRUCTURES

JACEK WOŹNIAK, PIOTR ZASKÓRSKI

Department of Organization and Management, Faculty of Cybernetics, Military University of Technology in Warsaw (*WAT*)

The article is devoted to the problems of the security management for the dispersed organizational structures. The security is one of the system criteria that determine the factual value of the organization, what tends to expose the information and decision-making processes in an article. These processes determine the efficiency of modern business units and directly affect the level of the overall security of dispersed organizations. The security of an organization can be seen in various aspects ranging from the system security (general, global) through the economic security to the information security. The process nature of the relationships between the links of an dispersed structure indicates the need for the creation of the operational security. A particular role is played by the ICT as a strengthening factor for the widely understood potential of the company and its economic security in this area, e.g. in the framework of stimulating processes of a group learning, using of dispersed databases, or verifying generated knowledge in terms of the efficiency and certainty of the decision-making processes in dispersed structures. This article is a kind of an outline of the concept of the new *H-H type* structural solution (the hybrid of the hierarchical and heterarchical configurations), taking the criterion of an information and economic security into an account.

Keywords: Dispersed Structure, H-H Structure, Network, Security, ICT, X-Engineering, Cloud Computing, Knowledge

1. Introduction

The enterprises management is largely related to the implementation of the decision-making functions and to shaping the structural configurations in a way that allows for taking the criterion of their efficiency into an account, while maintaining the given boundary conditions [12]. A solution of a problem sat in a such way requires the use of modern information technology that is conducive to shape the rational dependencies in the spatially and information dispersed structures.

Making the decisions by dispersed units is laden with the large dynamics, resulting often from the need for joining new participants to the network structure and modifying the goals of the organizational system (the newly-configured network). Available and used information and communication technologies can effectively support the decision-making and information processes, resulting usually in shortening the time of each task. A modern business organization requires, first and foremost, the focus on the problem of the security management in dispersed structures and on creation of the value in an organization in terms of the system approach with a particular comply with the requirements of the information security (as one of the dimensions of the general and economic security).

The results of the literature research in an area of the enterprise security modeling lead to the concept of a new structural solution which is a hybrid of hierarchical and heterarchical organizations. Therefore, there are being looked for a substantive basics to the specification of a new structural configuration (*H-H type* configuration) in this study – in reference to the *sustainability paradigm* [6], engraving the role and importance of the operational (and partially the strategic) criterion of the economic security, and taking the changes in the technological environment into an account (in the context of the dynamics of decision-making functions within an enterprise).

2. Dispersed organizational structures and the *H-H type* structure

The security management in the dispersed structures requires an identification of security determinants. It cannot be, in fact, *a priori* said that the hierarchical (vertical) structures guarantee the security of an organization to a lesser extent (both at the micro- and mesoeconomic levels) than the flat structures. Depending on the situational context of the decision-making processes, the business organizations are able to maintain the competitive potential taking a hierarchical structure, or in other situations a network structure. It is rare, in fact, that one of the predetermined activity model is able to ensure the economic security, both in the short and the long time horizon, taking changes in the environment (primarily in the proximal environment) into an account.

The degree of flattening an organizational structure is not a source of increasing the degree of an immunity of an organization to the both kinds of threats: existing in the environment ones (business, demographic, infrastructural etc.), as well the internal organizational mechanisms. Network structures are characterized by a high degree of operational flexibility and are focused on the efficient knowledge management (usually the criteria for selecting a business participant are knowledge and experience) [19]. However, it cannot be assumed that these are the key and sufficient factors to ensure the economic security. Network structures constitute the sources of additional threats to the organizations (the system) that are less clear to observe in the hierarchical structures. In addition, it should not be claimed that the hierarchical structures are the "outdated" solutions, that in the current business environment does not meet the criterion of an efficiency. It must also be noted that the integration of the hierarchical and heterarchical configurations should not affect negatively the degree of the use of the available IT solutions, and it may even increase that degree. It creates a base not only to integrate spatially and thematically dispersed information resources (databases) of co-participants, to use the Cloud Computing technology, or to use the Data Mining tools in a dispersed environment – but also (if not primarily) to verify the degree of usability, reliability and quality of created information and knowledge in an organizational system (e.g. through the implementation of the hierarchical control mechanisms).

The model of the *H-H structure* is developed for the specific circumstances and the specifics of the organizations' approach to achieving the objectives by the same company. Under certain conditions, the *H-H structure* can be successfully adapted and states the starting point for efficient business activities. The *H-H structure* differs from other structural configurations in the framework of the approach to exploit the potential of ICT solutions, mainly for the creation of the market position of an organization with the use of the Cloud Computing technology and Integrated Management Information Systems (IMIS). The primary benefit of an integration of hierarchical and heterarchical configurations, on the side of the first one, is to support processes of control and automatic switching to the mode of "one decision-making unit" in a situation of the external or internal (i.e. the business-natured) threats. At this point, it is observed the activation of the *Central Decision-making Unit* (CDU), as well the implementation of the tasks that require focusing the decision-making processes in one unit, or the need for the reallocation of the production capacities [30]. The network structures, in turn, give the opportunity, first of all, to free shaping the dependencies on the teams, e.g. in terms of choosing business participants, taking the criterion of knowledge into an account, and focusing on the creation of new knowledge in the system. There is visible the temporal nature of this structural type's functioning and of its openness on the communication with the environment, based on the platform of the Internet in this case [5, 13,

19, 21]. Therefore, the *H-H type* structure is able to exploit the potential of dispersed information resource management with an exposition of an information security. It can be assumed that the *H-H structure* is a specific class of a learning organization with a strong support of the planning and control functions.

3. The concept of X-Engineering and integration of dispersed structures

A reflection of the systemic perception of the security management processes in the spatially and information dispersed organizations is the concept of X-Engineering. Its main determinant is based on such functional assumptions of an organization, that will make it easier to connect to a multi-faced business relationships by a given business unit [3, 29]. This is a management concept adapted to the organizational changes at the micro- and mesoeconomic abstraction level. Narrowing the analysis spectrum solely to the environment of a single organization (also a spatially dispersed one) it may be noticed a necessity to implement the market relationships in a set of the internal customers (according to the process management model [29]). Turning on the mesoeconomic area, the special role is played by the relationships between co-participants in the network [29]. Therefore, it can be assumed that X-Engineering strategy is the premise for the transition from the hierarchical organizational model, the flat heterarchical model etc. into the *H-H model*. The particular importance in this concept is ascribed to the processes of enlarging the business activities by the possibility and potential of the external units, also with a regard to the implementation of planning and control functions. The X-Engineering triangle assumes, in fact, an integration and development of areas such as: processes, proposals and participations – with the proviso that participations of the external units, the extent of their interference in the organization's environment, as well their contribution in the form of information and knowledge resources is precisely defined [3]. In such a situation, it is difficult to talk about a balance (equivalence) of all the elements that make up a dispersed structure. The effect of primacy always occurs in this case. This phenomenon can be thought of as a permanent feature of the heterarchical structures [17, 20].

The concept of X-Engineering exposes the need for an operational integration of the activities conducted in an organization on the base of information and communication technologies [3, 29]. Bearing in mind the ICT solutions, there should be noted a special role of the transactional IMIS class systems, and the Business Intelligence systems (including the Data Mining systems). However, it should be noted that one of the ICT infrastructure's element, responsible for the integration of the dispersed structures, is an environment of the Internet [8]. Contemporary trends in the development of the IMIS suggest a raise in a significance of the Internet environment as a basic platform for the information exchange between processes'

executors in an organization. The IMIS are equipped with advanced analytical modules and adapted to carry out the process reengineering [29].

The examples of the evolution of the On-line transaction processing class systems (OLTP) can be both systems supporting manufacturing processes, such as the ERM class systems, e.g. EERP (*Extended Enterprise Resource Planning*), eERP (*electronic Enterprise Resource Planning*), @ERP (*active Enterprise Resource Planning*) [29], as well the DEM class systems (*Dynamic Enterprise Modeling*) [29], or the systems supporting the Customer Relationship Management (CRM) – which reflect the possibility/potential of Web 2.0 technologies, e.g. PRM (*Partner Relationship Management*), SRM (*Suplier Relationship Management*), VRM (*Visitors Relationship Management*), eCRM (*electronic CRM*), mCRM (*mobile CRM*) and ERM (*Employee Relationship Management*) [2, 16].

It is worth to note that not only the OLTP class systems support the security management in the dispersed structures. Among the Business Intelligence class systems (the analytical processing), there can be indicated e.g. the dedicated analytical devices, the DPA (*Decision Process Automation*) and IPA systems (*Intelligent Process Automation*). The IPA tools are also common termed as the BAM systems (*Business Activity Monitoring*), which differ from the classic Business Intelligence systems. The BAM systems, in turn, represent the core of systems based on the Balanced Scorecard (EPM – *Enterprise Performance Management*). These systems are defined as the BPM class systems (*Business Performance Management*) or CPM class systems (*Corporate Performance Management*) [7]. There is also worth to note the possibility of the implementation of the Corporate Business Intelligence Portals at this moment. There should also be highlighted the e-Business Intelligence systems and the EIP systems (*Enterprise Information Portal*), which are called the real-time BI systems in this case. There should also be mentioned the *enhanced Data Warehouse* (eDW), which gathers data from Internet and filters it into the Data Warehouse, and also searches for data by DW mechanisms, as well the *enhanced Knowledge Warehouse* (eKW) [7]. Not without a significance is also the cloud processing technology. The evolution's directions of the IMIS class systems and their impact on the economic security management in the dispersed structure are presented at the figure 1.

The application of the concept of X-Engineering as an integrating element of the area of an economic security management in the dispersed business organizations, as well the integration of the decision-making processes, and the processes of knowledge creation imply the information asymmetry phenomenon [28, 29]. This is the attribute that fits the foundations of the *H-H model* of a structural configuration and simultaneously is a response to an increase in the importance of the security of the information resources as an important component of the overall security of an organization. It should be noted, however, that the provision of an

information security is not a sufficient condition to ensure an economic security. The complementary factor is also the continuity of implementation basic, auxiliary and management processes (mainly in terms of creating a network of subcontractors and functioning of the internal customer mechanism) [9, 31]. On the other hand, it must be remembered that the maximization of an economic profit function does not state *ex definitione* an economic security.

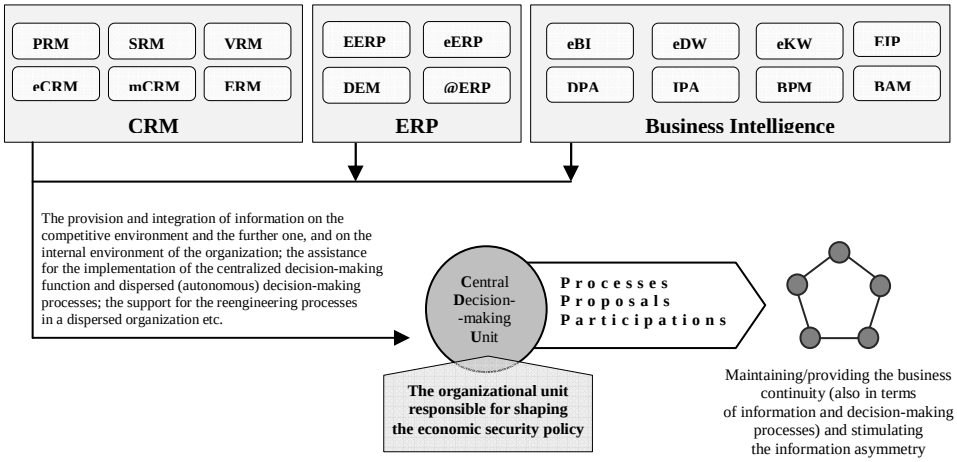

Figure 1. The directions for the evolution of the IMIS class systems and their expected impact on the security management in a dispersed structure. Source: own preparation

4. Information security in creation of the potential of business organizations

An information security as one of the dimensions of general and economic security, affects to a large extent on the state of the business continuity of the organizational system [31]. An information security is largely dependent on an ICT infrastructure used in a dispersed organization, as well on the nature of the managerial solutions, with a particular regard to the human factor and the organizational culture, and more concretely – the value system, the degree of confidence in a team, the activities geared for a cooperation and the criterion of knowledge [6, 31]. In general, the sources of an information security, especially in a virtual environment, are dichotomous-natured, i.e. determined by factors of the trust and the applied IT technology [23].

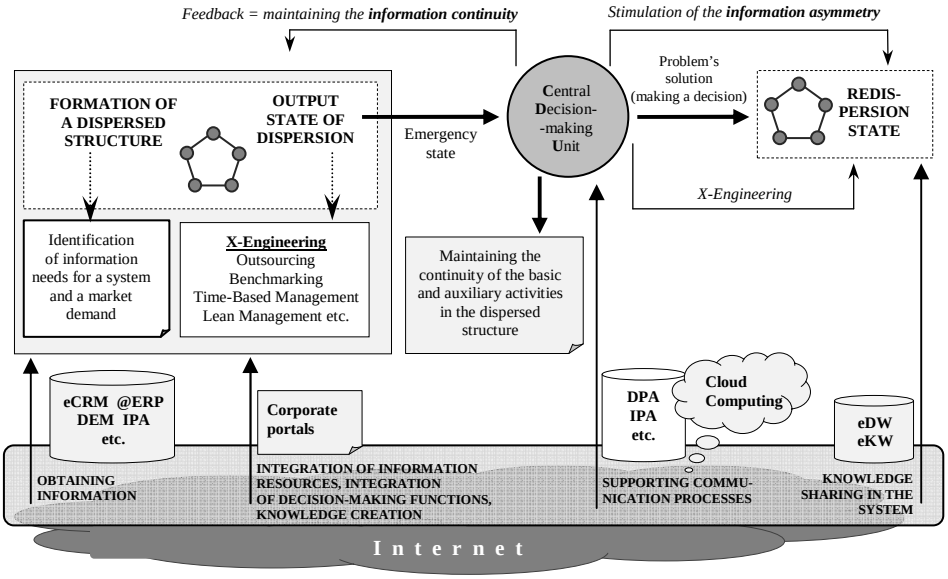

Figure 2. The emplacement of the X-Engineering concept in the information and economic security creation in a *H-H type* organization. Source: own preparation

An information resources management in a dispersed structure boils down to a skillful use of the potential of the Internet environment (and also intranet and extranet). Thanks to the possibility of obtaining, processing, sharing, sending and collecting data in a virtual environment, there is a base for the integration of data and information resources in a case of information security management. The basic technologies applied in an information security area are e.g. systems as follows: eCRM, @ERP, DEM, IPA, and the corporate portals, the CAx modules, the Cloud Computing technology, as well the eDW and the eKW (fig. 2). In this case, an information security comes down to providing the valuable resources (including new knowledge) to the units in a dispersed structure. The phenomenon of a knowledge creation has also the links to the feedback mechanism (stimulated by the CDU), mainly in the field of the correctness and suitability of created information, and the quality of data obtained by the various participants (fig. 2).

Each time movement of the *H-H structure* to a new state of dispersion requires setting up the assumptions (or bringing them up-to-date) for the procedures of an information asymmetry, e.g. about an access of the end users to data and information [28, 31]. It should also be noted that the CDU is primarily responsible for the security management of analytical data – and for the security of transactional data should be responsible the business units that make up a dispersed structure, as well the CDU in the emergency sates of the business activities.

5. Cloud computing in strengthening market position

The needs for a reorganization of a dispersed structure and a redefinition of the system[1] and elementary[2] usability functions (fig. 3) does not need to mean the organizational and technological difficulties. The one of the possible solutions is to apply the technology of the Cloud Computing (CC), which reduces the need for an expansion of each organization's ICT infrastructure, and there can even be acceptable (as well advisable) an abandonment of that kind of infrastructure in managing information and economic security in a short- and long-time horizon [1, 11, 22]. To the basic services in the Cloud Computing technology are included: SaaS (*Software-as-a-Service*) – an access to the various applications, e.g. an *online* archive, an email account, or spreadsheets and Web forms, IaaS (*Infrastructure-as-a-Service*) – an access to the *storage*-type infrastructure, PaaS (*Platform-as-a-Service*) – an access to the entire platforms for applications' development, i.e. servers, databases etc., and CaaS (*Communication-as-a-Service*) – an access to the communication platforms) [1, 15, 24].

The CC technology is focused on minimizing the costs of operational activities of a dispersed structure, e.g. in terms of shortening the communication time between the co-participants, reducing the costs of maintaining and developing an infrastructure, the data processing [22], increasing the enterprises' flexibility [1, 15, 26, 27] (and, thus, of the whole organizational system in terms of regarding the data and information processing in the Internet environment [27]). The CC technology allows also to increase the performance and efficiency with limited production capacities and the time to react to the occurrences identified in the environment of the private, public, social, or hybrid clouds [1, 15]. Therefore, it can be assumed that Cloud Computing is a technology which advancements the strengthening of the market position of an enterprise, as well an entire dispersed structure, e.g. in terms of supporting the Business Continuity Management and increasing the efficiency of knowledge creation. The Cloud Computing gives the ability to scale/diversify the products or activities, according to the demand expressed by the customers [1, 14], facilitates the collaboration within the spatially dispersed structures, and increases the flexibility of the data processing in the cloud.

[1] Marked at fig. 3 by the symbol $U_{syst}(t)$ – as a time function.
[2] Marked at fig. 3 by the symbol $u_n(t)$ for each element/participant in the network structure (spatially and information dispersed) – as a time function.

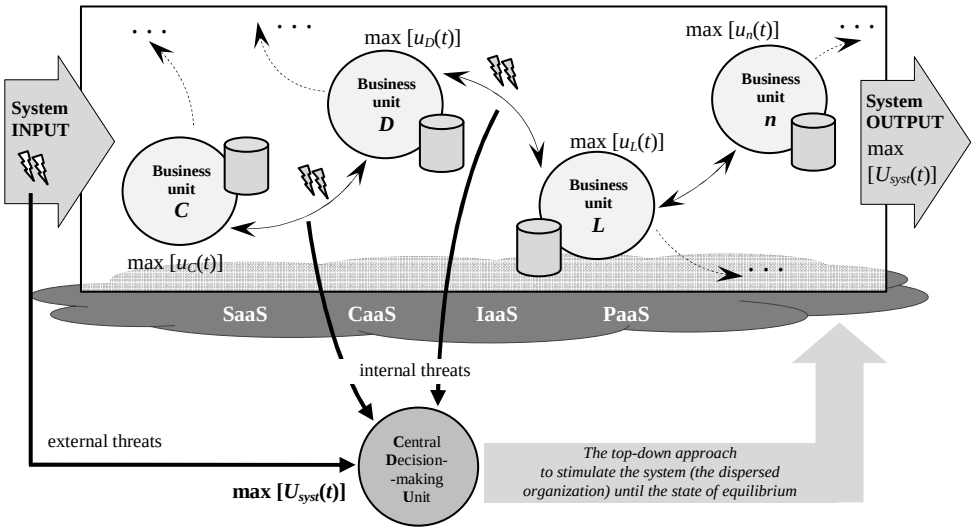

Figure 3. The Cloud Computing technology in strengthening the market position (by maximizing the utility functions) of the spatially and information dispersed *H-H structure*. Source: own preparation

The CC technology is a specific link between the needs of the customers (the internal and external ones), as well the other enterprises. Therefore, this technology is applied in the processes of maximizing the utility functions of different classes of stakeholders in dispersed structures (fig. 3). An important role in strengthening

the market position of a spatially and information dispersed organization is the CDU, which directly uses the Cloud Computing technology, applying it only to communicate with the co-participants in the framework of cooperation proposals' notification, as well with the external customers (as an optional feature of the CDU in the *H-H type* structure). The CC is primarily a technology dedicated to dispersed structures, supporting the operation of the elementary business units, e.g. in an area of the use of professional solutions of the IMIS class as a CC service. The CDU in a situation of emergency (internal- or external-natured) takes over the decision-making function and reorganizes the triad: processes, proposals, participations (as already mentioned), using e.g. the Expert Systems, Executive Information Systems, Artificial Intelligence Systems, or Data Mining systems.

The Central Decision-making Unit stimulates dispersed business units (participants) to take new actions (improving-, mending- and innovating-natured ones), and indirectly it specifies a new data processing rules based on the potential and possibilities of the CC technology. The individual business units continue to have an access to their own databases (e.g. in the form of the information "islands") in a state of emergency, what gives them the ability to maximize their own utility function – what also indirectly affects the enhancement of the security and the market position of the entire network structure. What more, the Cloud Computing technology effects in concentrating by an organization on its core business activities (i.e. their specialization), using mechanisms of the economies of scale (through an increase in a productivity), as well as increasing the transparency of the system operation/organization, and in supporting the processes of business activities' controlling [10]. It is also worth to notice that the flexible and efficient implementation of business projects in a dispersed environment, e.g. in network or virtual structures, requires the use of Web applications and services[3] [24, 25].

The computing in a cloud in a spatially and information dispersed organization is connected with formation of the specific risks, associated e.g. with the necessity of an additional geographic dispersion of services and data stored in a cloud, as well with ensuring the security of confidential data collected and processed on the external servers [4, 25]. However, while a data loss is not a threat to the enterprise (by SLA agreements guaranteeing the time of services' functioning at 99.999%, as well by making the backups [26]), there may appear a risk of the temporary loss of access to the important classes of operational data in that kind of situation. Making a decision with the use of the CC technology's potential in managing an economic security of a dispersed organization by stimulating the competitive potential might also be a major source of the risks (mostly of the decision-making nature, while not necessarily of the technological one), of which an organizational system (and especially the CDU) should be aware.

[3] For example the Web Services standards, using e.g. the Cloud Computing technology and SOA (*Service Oriented Architecture*).

6. Conclusions

The network structures are neither a *sine qua non* nor a sufficient condition for ensuring an economic security in an operational mode in a dispersed environment. Thus, there is a need to integrate the hierarchical and heterarchical (H-H) structures what represents a new approach to provide and maintain the desired level of a security of an organisation at both a micro- and mesoeconomic levels. Providing and maintaining an economic security in spatially and information dispersed structures should be identified with actions aimed at adapting to changes and proactive stimulating desired changes in an organization and their environment – in order to maximize the elementary and system utility functions. Therefore, the role of the CDU is special and important from the perspective of ensuring and maintaining the desired level of an economic security.

It should also be added that the main dimension of an economic security is not only the continuity of the basic and auxiliary processes (that generate a value in a dispersed system), but also the continuity of information and decision-making processes, with a particular regard to the information asymmetry phenomenon in the area of transactional and analytical data management, as well the creation and verification of knowledge in an organization. An economic security can be seen through the prism of the learning processes in a business unit (also in a network structure), not only on a base of their own (individual/single) experiences, but above all on a base of the multidimensional operational interactions with entities in an environment (according to the assumptions of the X-Engineering concept). An organization's security is also determined by the degree of its openness to the potential of the external units.

Not without a significance is also the fact that an economic security management requires the cybernetic frame. The dispersion state (i.e. the state of the secure and stable implementation of an economic security policy, including the areas of creating the added value and maximizing the utility functions) cannot fully meet the complexity and instability of both a dispersed structure and its environment. Therefore, the justifiable activities seem to be the specification and description of the *controlled system* (a network of the business units) and the *controlling system* (in a form of the Central Decision-making Unit). This dual nature of the decision-making functions is an attempt to increase the operational flexibility of a dispersed structure under the conditions created by ICT, notably in the field of a communication in the Internet environment (the Cloud Computing), as well of the processing and integration of transactional and analytical data. It should be mentioned, however, that currently available and developed ICT technology can be an important source of threats in ensuring and maintaining an economic security for an entire dispersed system, as well for its individual components.

The article is financed from the public science funds as the research project No. RMN 705/2012 (Faculty of Cybernetics, Military University of Technology in Warsaw), entitled "Security Processes' Modeling" (in Polish: "Modelowanie procesów bezpieczeństwa").

REFERENCES

[1] Armbrust M. et al. (2010) *A View of Cloud Computing*, Communications of The ACM, 4(53)/2010, pp. 50-58.

[2] Buchnowska D. (2010) *Systemy CRM*, Wrycza S. [ed.]: *Informatyka ekonomiczna*, Polskie Wydawnictwo Ekonomiczne, Warszawa, Poland, pp. 373-403.

[3] Champy J. (2003) *X-engineering przedsiębiorstwa*, Placet, Warszawa, Poland.

[4] Chen Y., Paxson V., Katz R.H. (2010) *What's New About Cloud Computing Security?*, University of California at Berkeley, Technical Report No. UCB/EECS-2010-5, http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-5.html

[5] Czakon W. (2012) *Sieci w zarządzaniu strategicznym*, Wolters Kluwer, Warszawa, Poland.

[6] Grudzewski W.M., Hejduk I.K. (2011) *Przedsiębiorstwo przyszłości. Zmiany paradygmatów zarządzania*, Master of Business Administration, 1(108)/2011, pp. 95-111.

[7] Januszewski A. (2008) *Funkcjonalność informatycznych systemów Zarządzania. T.2. Systemy Business Intelligence*, Wydawnictwo Naukowe PWN, Warszawa, Poland.

[8] Jaruga A. (2010) *Technologia teleinformatyczna w organizacji wirtualnej*, Wydawnictwo Politechniki Poznańskiej, Poznań, Poland.

[9] Kaczmarek T.T., Ćwiek G. (2009) *Ryzyko kryzysu a ciągłość działania: Business Continuity Management*, Difin, Warszawa, Poland.

[10] Kaiserswerth M. et al. (2012) *White Paper: Cloud Computing*, Schweizerische Akademie der Technischen Wissenschaften, Zürich, Swiss, https://www.satw.ethz.ch/organisation/tpf/tpf_ict/box_feeder/2012-11-06_2_SATW_White_Paper_Cloud_Computing_EN.pdf

[11] Koehler Ph., Anandasivam A., Ma D. (2010) *Cloud Services from a Consumer Perspective*, Proceedings of the Sixteenth Americas Conference on Information Systems, Lima, Peru, August 12-15, 2010.

[12] Kuciński K. (2010), *Przedmiot nauk ekonomicznych*, Kuciński K. [ed.]: Metodologia nauk ekonomicznych. Dylematy i wyzwania, Difin, Warszawa, Poland, pp. 53-83.

[13] Łobejko S. (2010) *Przedsiębiorstwo sieciowe: zmiany uwarunkowań i strategii w XXI wieku*, Szkoła Główna Handlowa, Warszawa, Poland.

[14] Marston S. et al. (2011) *Cloud computing – The business perspective*, Decision Support Systems, 51/2011, pp. 176-189.

[15] Mell P., Grance T. (2011) *The NIST Definition of Cloud Computing (Draft)*, NIST Special Publication 800-145 (Draft), National Institute of Standards and Technology, Gaithersburg, USA.

[16] Mazur D. (2010) *Systemy informatyczne zarządzania relacjami z klientami*, Zawiła-Niedźwiecki J., Rostek K., Gąsiorkiewicz A. [eds.]: Informatyka gospodarcza. T.3, C.H. Beck, Warszawa, Poland, pp. 653-672.

[17] Mrożek A. (2013) *Zarządzanie bezpieczeństwem organizacji o strukturze heterarchicznej*, http://bezpieczna.uek.krakow.pl/artykol.pdf

[18] Perechuda K. (2007) *Dyfuzja wiedzy w przedsiębiorstwie sieciowym: wizualizacja i kompozycja*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław, Poland.

[19] Płoszajski P. (2013) *Organizacja przyszłości: przerażony kameleon*, http://www.allternet.most.org.pl/SOD/Heterarchia%20prof._Ploszajski_-_Organizacja_przyszlosci.pdf

[20] Polańska K. (2013) *Ewolucja przedsiębiorstwa w środowisku wirtualnym*, Kwartalnik Nauk o Przedsiębiorstwie, 1(26)/2013, pp. 30-43.

[21] Report: *Cloud Computing and Sustainability: The Environmental Benefits of Moving to the Cloud* (2010) Accenture.

[22] Sankowska A. (2009) *Organizacja wirtualna: koncepcja i jej wpływ na innowacyjność*, Wydawnictwa Akademickie i Profesjonalne, Warszawa, Poland.

[23] Satyanarayana S. (2012) *Cloud Computing: SaaS*, GESJ: Computer Science and Telecommunications, 4(36)/2012, pp. 76-79.

[24] Szermanowicz K. (2012) *Wpływ możliwości komunikacyjnych systemów IT na rozwój przedsiębiorstw sieciowych*, Łobejko S. [ed.]: Przedsiębiorstwo sieciowe i inne formy współpracy sieciowej, Szkoła Główna Handlowa, Warszawa, Poland, pp. 159-185.

[25] Tan C., Liu K., Sun L. (2013) *A design of evaluation method for SaaS in Cloud Computing*, Journal of Industrial Engineering and Management, 6(1)/2013, pp. 50-72.

[26] Urban W. (2009) *Narzędzia służące zwiększaniu elastyczności*, Kasiewicz S. et al.: Metody osiągania elastyczności przedsiębiorstw: od zarządzania zasobowego do procesowego, Szkoła Głowna Handlowa, Warszawa, Poland, pp. 117-148.

[27] Woźniak J., Zaskórski P. (2009) *Asymetria informacyjna w zarządzaniu bezpieczeństwem organizacji procesowych*, Nowoczesne Systemy Zarządzania, 4/2009, pp. 187-198.

[28] Zaskórski P. (2012) *Asymetria informacyjna w zarządzaniu procesami*, Wojskowa Akademia Techniczna, Warszawa, Poland.

[29] Zaskórski P., Gonciarski W. (2011) *Struktury i strategie zarządzania organizacją*, Zaskórski P. [ed.]: Zarządzanie organizacją w warunkach ryzyka utraty informacyjnej ciągłości działania, Wojskowa Akademia Techniczna, Warszawa, Poland, pp. 11-46.

[30] Zaskórski P., Woźniak J. (2009) *Ciągłość informacyjno-decyzyjna warunkiem bezpieczeństwa organizacji gospodarczej*, Gonciarski W., Zaskórski P. [eds.]: Wybrane koncepcje i metody zarządzania początku XXI wieku, Wojskowa Akademia Techniczna, Warszawa, Poland, pp. 133-145.

INFORMATION SYSTEMS IN MANAGEMENT

Information Systems in Management (2014) Vol. 3 (1) 77−88

ON USING DATA MINING TECHNIQUES FOR CONTEXT-AWARE STUDENT GROUPING IN E-LEARNING SYSTEMS

DANUTA ZAKRZEWSKA

Institute of Information Technology, Lodz University of Technology

Performance of an e-learning system depends on an extent to which it is adjusted to student needs. Priorities of the last ones may differ in accordance with the context of use of an e-learning environment. For personalized e-learning system based on student groups, different distribution of the groups should be taken into account. In the paper, using of data mining techniques for building student groups depending on the context of the system use is considered. As the main technique unsupervised classification is examined,. Context parameters depending on courses and student models are tested. Experiment results for real student data are discussed.

Keywords: E-learning, Data Mining, Student Grouping, Context-Awareness

1. Introduction

In e-learning good student performance depends on en extent to which educational environment is tailored to learners' profile [1]. Grouping students of similar characteristics enables to adjust educational system to groups of colleagues, who should learn together from the same resources. However features of educational environment should also differ depending on the context of the system use. Context aware differentiation of teaching materials seems to be an important feature of an e-learning system [2]. The research aims at examining data mining techniques namely unsupervised classification methods for student groups' creating, in different context of the educational system usage.

The paper is organized as follows. In the next section, literature review concerning context-aware personalized e-learning systems, student grouping as well as application of data mining in e-learning will be presented**.** Then, context aware models in e-learning systems will be described. The following section will be devoted to application of clustering techniques for student grouping taking into account context of use.

Finally, the case study of student profiles characterized by learning styles will be considered and experiments on building context aware groups for real students data will be described and discussed. The paper will be finished with concluding remarks and future research presentation.

2. Related work

Data mining techniques were very often applied as personalization tools in e-learning systems. The broad review of the research in that area can be found in [3, 4]. Cluster analysis was used to group students on the basis of their behaviors (see [5, 6]) or individual traits (see [4] for example). Many authors connected clustering with different data mining techniques to increase efficiency of obtained results. Shen et al. [7] applied cluster analysis together with sequential pattern mining to group students according to their learning activities. Tang and McCalla [8], in turn, integrated clustering technique together with collaborative filtering for building recommendations of course contents. The last technique combined with association rules mining was often used for building student recommendations (compare [9, 10]).

In e-learning systems, student groups were often built for recommendation purposes. Authors grouped students taking into account their behaviors, pages they visited or historical navigational paths ([6, 8]), as well as learner cognitive styles or usability preferences [4].

Many of the researchers emphasized an importance of a context in a personalization process (see [2, 11, 12], for example). The broad review of context parameters as well as context aware e-learning systems was presented in [13]. Context-awareness was very often considered for recommendation purposes. Andronico et al. [14] built multi-agent system to suggest students educational materials taking into account learners' behavior and their preferences while using different mobile devices. Rosaci and Sarné, in turn, considered both: student's profile and an exploited device [15]. Their recommendations were built on the basis of the time spent by student on the particular Web site, taking into account type of a device used for navigating. Zaïane [16] proposed an agent, which aims at recommending learning activities or shortcuts in course web-sites, considering learners' historical activities. Using of Naïve-Bayes models for building context-aware group recommendation was proposed by Zakrzewska [17].

3. Context-aware models in e-learning

Dey [18] defined context as "any information that can be used to characterise the situation of an entity". Context of use plays an important role in e-learning. Students' needs may differ depending on the situation of use and what is more, different student features may be important in the cases of different courses.

Dey [18] said that "the system is context-aware if it uses context to provide relevant information or services to the user, where relevancy depends on the user's task". In educational systems context-awareness should be taken into account in order to obtain its personalization features. In that case, relevant information can have the form of teaching materials tailored into learners' needs according to the usage situation.

Das et al. [19] distinguished three types of context parameters in e-learning: personal, abstraction and situation. The first one is connected with student personal information, personality type and the level of expertise. Situation context describes learner situation, network and device he uses [19]. In our considerations we will focus on abstraction context which concerns the information of student preferences and learning styles, in the situation of the course, that learner attends. We will assume that the context model is different depending on the course that student enrolls on.

Let us assume that each student *ST* is described by *N* attributes, which may indicate their learning styles or other preferences. A tuple *ST* of a student is of the form:

$$ST = (st_1, st_2, ..., st_N), \quad st_i \in DOM(S_i), \tag{1}$$

where $DOM(S_i)$ stands for the domain of S_i .Further, we will assume, that the attributes are of different importance, depending on the context of an educational system usage. Let the context *CN* will be described by *N* weight parameters:

$$CN = (cn_1, cn_2, ..., cn_N), \tag{2}$$

where $cn_i \geq 0$, $i=1,2,...,N$; mean the importance of the *i*-th attribute of the student model in the context *CN*. We will also suppose that

$$\sum_{i=1}^{N} cn_i = 1, \quad cn_i \geq 0, \quad i = 1,..., N. \tag{3}$$

If any of the attributes has no significance in the considered context, respective weight value is equal to 0.

To take into account the context of use, we will include context vector into the grouping process. As the most important, student features according to the biggest weights of the context vector will be considered. Cluster analysis of students' data was also broadly examined in [4], where different algorithms were considered in

order to build groups of students of similar needs. Investigations, presented there, showed advantages of unsupervised classification. The main problem, which arises in the current research, consists in including context into cluster analysis tool.

Assuming that each course can be modeled by a vector of weights related to respective student attributes, we propose to present context-aware grouping as clustering problem with weighted distance function. Then, similarity of group members will be measured by a function, where contribution of each of the attributes depends on the respective weight values.

Let x and y be vectors $x = (x_1,x_2,...,x_N)$ and $y = (y_1,y_2,...,y_N)$ and let $w = (w_1,w_2,...,w_N)$ denote a vector of weights, where w_i is non-negative for every $I = 1,..,N$ and fulfils (3). Then a weighted distance function d_w will take the form:

$$d_w(x, y) = \sum_{i=1}^{N} w_i d(x_i, y_i) \tag{4}$$

Such way of including weights into a distance function will not change none of its metric properties and will enable taking into account priorities of each of the attributes. However according to such approach only distance based clustering can be applied, but the results of the proposed technique should not depend on the choice of an algorithm.

Obtained student groups should be different depending on the context and their quality should be measured by taking into account the most important attributes, from the context point of view. Students from the same group should have the most similar features, which tutors decide as the crucial in the considered context. The cluster quality can be examined by calculating standard deviations separately for each attributes within clusters. The smallest value should concern features of the biggest importance.

4. Learning styles and usability preferences case studies

Let us consider student models based on their dominant learning styles. We will examine the model which was often used to ensure adaptivity features of e-learning systems [20], proposed by Felder and Silverman [21]. It is based on "Index of Learning Style" (ILS) questionnaire developed by Felder and Soloman [22]. The results of ILS questionnaire indicate preferences for 4 dimensions from among excluding pairs: *active* vs. *reflective*, *sensing* vs. *intuitive*, *visual* vs. *verbal*, *sequential* vs. *global*. The index, obtained by each student, has the form of the odd integer from the interval [−11; 11], assigned for all the dimensions.

Thus student learning style model *SL* is represented by 4 integer attributes:

$$SL = (sl_1, sl_2, sl_3, sl_4) = (l_{ar}, l_{si}, l_{vv}, l_{sg}) . \quad (5)$$

Element l_{ar} means scoring for *active* (if it has negative value) or *reflective* (if it is positive) learning style, and respectively l_{si}, l_{vv}, l_{sg} are scores for all the other dimensions, negative values are in cases of *sensing*, *visual* or *sequential* learning styles, while positive values are in cases of *intuitive*, *verbal* or *global* dominant learning styles.

That way $N = 4$. We will consider building student groups in the context of different courses, then each course *CR* will be modeled by a vector of weights, which signify the importance of different learning style preference for the course:

$$CN = CR = (w_{ar}, w_{si}, w_{vv}, w_{sg}) . \quad (6)$$

As the second student model usability preferences will be considered. As the most important design categories, deciding on Web sites usability, which should be evaluated by users, Marsico and Levialdi [23] mentioned information representation and appearance, access, navigation and orientation as well as the informative content architecture of the sites. Investigations presented in [24] showed that students put special attention to graphical attractiveness of Web sites and the efficiency which means a short time of loading the sites. Students also emphasized the importance of advanced search possibilities. Consequently, five preferences for portal features are taking into account: informative contents, graphics, navigation, efficiency and search possibilities. Students were asked to score the importance of each of the feature, assigning from 1 to 5 scores. Values equal to 1 or 2 mean that a student does not put attention to the portal characteristic, 3 means that a learner does not distinguish the importance of considered feature from among the others, finally values 4 or 5 mean that the usability trait is important for the student. Let *SU* denote student usability preference model. Taking into account the meaning of the score values, *SU* is represented by 5 attributes:

$$SU = (su_1, su_2, su_3, su_4, su_5) , \quad (7)$$

where su_1 means scoring for importance of informative content, su_2 scoring for importance of graphics, su_3 scoring for importance of navigation, su_4 scoring for importance of efficiency of the system and finally su_5 means scoring for the importance of search possibilities. Then $N = 5$ and consequently, the respectful course model of weights is of the form:

$$CR = (wu_1, wu_2, wu_3, wu_4, wu_5). \quad (8)$$

5. Experiment results and discussion

Experiments aimed at checking, how including context into clustering process will change student group characteristics. Two main attribute categories were considered: dominant learning styles and usability preferences. The research was done on the basis of experiments conducted on the two trial sets of students from Technical University of Lodz: the set A of 22 learners studying the same master course of Information Systems in Management, and the set B of 56 part-time Computer Science students, who were also graduates of other programs. Firstly, students filled ILS questionnaire and answered questions concerning usability needs then groups of similar preferences were built, taking into account each course context. Finally grouping effects were evaluated.

For the grouping purpose 8 different courses were considered. During the courses: *CR1, CR2, CR3, CR4* students were grouped according to their dominant learning styles. For the courses *CR5, CR6, CR7, CR8,* in turn, usability preferences were taken into account. All of them characterised by different weight vectors.

It was decided that a student model for the course *CR1* is global and all the learning styles attributes are equally important, while during the preparation of the course *CR2* only dimensions: *active/reflective* or *visual/verbal* should be taken into account. This dimension has the highest priorities in both of the courses: *CR3* and *CR4,* however its importance is much bigger for the course *CR4*. In this course the dimension *sensing/intuitive* is not considered. Learning style dimensions' weights for all the courses are presented in Table 1.

Table 1. Weights for courses. Case of dominant learning styles

Course	w_{ar}	w_{si}	w_{vv}	w_{sg}
CR1	1/4	1/4	1/4	1/4
CR2	1/2	0	1/2	0
CR3	1/6	1/6	1/2	1/6
CR4	1/20	0	9/10	1/20

Models are global in the case of the course *CR5* and all the usability preferences attributes are of the same priority. Educational materials prepared for the course *CR6* were distinguished depending on three usability preferences: informative content, graphics and navigation, not taking into account efficiency and search possibilities. Informative content is of the highest priority in the course *CR7*, while all the others attributes are of the same importance. For the course *CR8,* navigation is the most important feature, informative content is the second one, all the other features are of the same priorities. Weights of usability preferences for all the courses are presented in Table 2.

During experiments students' data from the sets A and B were clustered, taking into account all the courses' needs. Context of the course was included as weight values into distance function. Students were divided to 5 groups, the number for which clustering schemas was stated to be optimal in many cases, while student grouping [25].

Table 2. Weights for courses. Case of usability preferences

Course	wu_1	wu_2	wu_3	wu_4	wu_5
CR5	1/5	1/5	1/5	1/5	1/5
CR6	1/3	1/3	1/3	0	0
CR7	1/3	1/6	1/6	1/6	1/6
CR8	1/5	1/10	1/2	1/10	1/10

Groups were built by using k-means algorithm implemented in the Open Source Weka software [26], taking into account Manhattan distance function.

During the process of quantitative analysis all the obtained clusters were compared, taking into account course context. Standard deviations within clusters were used to examine its qualities. Mean values of attributes decided of cluster profiles. The results for the set A and courses *CR1*, *CR2* and *CR4* are presented in Table 3 . Clusters obtained for the course *CR3* were of the same parameters as the ones obtained for the course *CR1*. Weights close or equal to 0 changed the structure of obtained groups. '-' means that the attribute was not taken into account during the clustering process.

To examine, how weights included into clustering process influence grouping effects, standard deviations for the whole sets and the most important attributes were calculated and compared: l_{ar} l_{vv} for the set A and courses *CR1, CR2, CR4* and l_{vv} for the set B and courses: *CR1, CR2* and *CR4*.

In the first case averages of standard deviations are respectively equal to 1.92, 1.70 and 3.23 for l_{ar}, and 2.87, 1.32 and 2.32 for l_{vv}. In the case of set B obtained values are respectively: 2.83, 2.15, 4.49 for l_{ar} and 2.7, 2.69 and 1.55 for l_{vv}. It can be easily noticed that removing not important attributes from the clustering process ameliorated the quality of obtained groups.

Table 4 contains respective values of means and standard deviations of attributes within clusters obtained for the set B and courses *CR1, CR2* and *CR4*. Similarly to the previous case the results got for the course *CR3* were the same as for *CR1*.

Table 3. Set A. Mean values of attributes within clusters

Crs	Clst. No	Inst.	Mean values				Standard deviations			
			l_{ar}	l_{si}	l_{vv}	l_{sg}	l_{ar}	l_{si}	l_{vv}	l_{sg}
CR1	1	4	-5	-5	-5	-4	1	1.93	1	4.76
	2	4	-6	-4	-9	-1	2.83	5.74	1	2
	3	4	6	-6	-10	-3	2.83	1.91	2.83	4.43
	4	2	2	-9	1	1	1.41	0	5.66	0
	5	8	-5	1	-3	3	2.07	2.25	3.85	1.41
CR2	1	8	-5	-	-5	-	1.07	-	2.83	-
	2	5	-7	-	-9	-	1.79	-	0	-
	3	5	5	-	-11	-	3.63	-	2.61	-
	4	1	3	-	5	-	0	-	0	-
	5	3	-1	-	-3	-	2	-	1.15	-
CR4	1	2	-4	-	-6	-8	1.41	-	1.41	1.41
	2	8	-6	-	-9	-1	1.49	-	1.69	3.20
	3	2	3	-	-10	-7	2.83	-	1.41	2.83
	4	2	-1	-	7.5	2	5.66	-	3.54	1.41
	5	8	-1	-	-5	1	4.78	-	3.54	1.28

Table 4. Set B. Mean values of attributes within clusters

Crs	Clst. No	Inst.	Mean values				Standard deviations			
			l_{ar}	l_{si}	l_{vv}	l_{sg}	l_{ar}	l_{si}	l_{vv}	l_{sg}
CR1	1	8	3	-11	-5	-5	2.39	2.33	2.83	4
	2	13	-7	-3	-1	-3	2.23	3.38	3.38	2.66
	3	12	-3	-6	-7	-5	2.58	3.13	1.51	3.57
	4	15	7	-7	-3	-1	3.40	2.56	3.64	2.33
	5	8	-9	7	-9	1	3.55	2.71	2.14	3.21
CR2	1	8	1	-	-3	-	1.51	-	2.14	-
	2	13	-7	-	-1	-	2.51	-	3.15	-
	3	4	7	-	-2	-	1.63	-	4.76	-
	4	12	2	-	-5	-	1.97	-	1.80	-
	5	19	-5	-	-9	-	3.12	-	1.61	-
CR4	1	12	1	-	-3	-3	5.52	-	0.98	1.71
	2	6	-5	-	3	-2	3.77	-	2.34	3.88
	3	13	-3	-	-7	-5	3.80	-	1.70	2.24
	4	13	-1	-	-5	-1	4.82	-	1.30	2.09
	5	12	-5	-	-9	1	4.55	-	1.44	2.97

Table 5 contains respective values of means and standard deviations of attributes within clusters obtained for the set A and courses *CR5* and *CR6*. Similar results for the set B are presented in Table 6. For both of the sets results obtained for the course *CR7* and *CR8* are the same as for the course *CR5*.

For the courses *CR5* and *CR6* averages of standard deviations of su_1 ,su_2,su_8, which are the most important in the course *CR6* were considered. In the case of the set A the averages for *CR5* are respectively equal to 0.1, 0.61, 0.29, while for *CR6* they take the value: 0, 0.61, 0.1. Similarly to the case of learning styles preferences one can conclude that removing of attributes of the low importance from the clustering process ameliorated the quality of obtained groups.

Table 5. Set A. Mean values and standard deviations of attributes within clusters

Crs	Cl. No	Inst.	Mean Values					Standard deviations				
			su_1	su_2	su_3	su_4	su_5	su_1	su_2	su_3	su_4	su_5
CR5	1	3	4	4	4	2	3	0	0.58	1	0.58	1.15
	2	4	5	3	4	5	4	0.5	0	0	0.5	0.5
	3	5	4	3	5	4	3	0	0.89	0.45	0.45	0.55
	4	3	5	5	4	4	3	0	0.58	0	0.58	0.58
	5	7	5	4	5	4	3	0	0.98	0	0.69	0.95
CR6	1	4	4	3	4	-	-	0	0.5	0.5	-	-
	2	6	5	5	4	-	-	0	0.98	0	-	-
	3	2	4	3	5	-	-	0	0	0	-	-
	4	7	5	4	5	-	-	0	0.97	0	-	-
	5	3	4	4	5	-	-	0	0.58	0	-	-

Table 6. Set B. Mean values and standard deviations of attributes within clusters

Crs	Cl. No	Inst.	Mean Values					Standard deviations				
			su_1	su_2	su_3	su_4	su_5	su_1	su_2	su_3	su_4	su_5
CR5	1	7	4	3	4	3	2	1	1.13	1.15	1.11	0.53
	2	22	4	4	4	4	4	0.58	0.66	0.70	0.73	0.77
	3	1	1	1	3	5	5	0	0	0	0	0
	4	17	4	3	5	5	4	0.66	0.87	0.24	0.94	0.79
	5	9	5	4	4	3	3	0.33	0.44	1.05	0.97	0.60
CR6	1	8	5	4	3	-	-	0.52	1.19	0.46	-	-
	2	16	4	4	4	-	-	0.65	0.72	0	-	-
	3	1	1	1	3	-	-	0	0	0	-	-
	4	21	4	4	5	-	-	0.54	0.85	0	-	-
	5	10	5	4	5	-	-	0	1.26	0	-	-

In the case of the set B, for the course *CR5* average standard deviations for first three attributes are equal respectively to 0.51, 0.62, 0.42.. For the course *CR6*, in turn, they are equal to 0.34, 0.80 and 0.12. Again, it could be easily noticed that removing of the attributes before the grouping process improved the quality of obtained clusters.

6. Conclusion

In the paper application of unsupervised classification for student groups' creating, in different context of the educational system usage, is investigated.

The context of the courses is considered, taking into account such student features as dominant learning styles and usability preferences. In the proposed method course context is presented as a vector of weights, included into clustering process. The effects are evaluated on the basis of the experiments done on the datasets of real students.

Tests showed that using of weights can ameliorate qualities of obtained groups if they differ significantly. What is more qualities of the biggest clusters were better from the most important attributes point of view.

Future research should consist on further experiments concerning more number of attributes of different meaning and importance as well as using different clustering technique. More precise cluster validation technique should be also considered.

REFERENCES

[1] Beaudoin M.F. (2002) *Learning or Lurking? Tracking the "Invisible" Online Student*, Internet & Higher Educ., 2/2002, 147-155.

[2] Schmidt A., Winterhalter C.(2004) *User Context Aware Delivery of E-learning Material: Approach and Architecture*, J. Univers. Comput. Sci.,10/2004, 38–46.

[3] Romero C., Ventura S. (2010) *Educational Data Mining: a Review of the State of the Art*, IEEE T. Systems, Man & Cybernetics, Part C: Applications & Reviews 6/2010, 601-618.

[4] Zakrzewska D. (2012) *Eksploracja danych w modelowaniu użytkowników edukacyjnych systemów internetowych*, AOW EXIT, Warszawa, Poland.

[5] Perera D., Kay J., Koprinska I., Yacef K., Zaïane O.R. (2009) *Clustering and Sequential Pattern Mining of Online Collaborative Learning Data*, IEEE T. Knowl. Data En., 6/2009, 759-772.

[6] Talavera L., Gaudioso E. (2004) *Mining Student Data to Characterize Similar Behavior Groups in Unstructured Collaboration Spaces*, Workshop on Artificial Intelligence in CSCL. 16th European Conference on Artificial Intelligence, 17–23.

[7] Shen R., Han P., Yang F., Yang Q., Huang J. (2003) *Data Mining and Case-based Reasoning for Distance Learning*, Journal of Distance Education Technologies, 3/2003, 46–58.

[8] Tang T., McCalla G. (2005) *Smart Recommendation for an Evolving E-learning System*, International Journal on E-Learning, 1/2005, 105–129.

[9] Minaei-Bidgoli B., Tan P., Punch W. (2004) *Mining Interesting Contrast Rules for a Web-based Educational System*, The Twenty-First International Conference on Machine Learning Applications, 1-8.

[10] Wang F. (2002) *On Using Data-Mining Technology for Browsing Log File Analysis in Asynchronous Learning Environment*, Conference on Educational Multimedia, Hypermedia and Telecommunications, 2005–2006.

[11] Jovanowić J., Gašewić D., Knight C., Richards G. (2007) *Ontologies for Effective Use of Context in E-learning Settings*, Educ. Technol. Soc., 10/2007, 47-59.

[12] Yang S.J.H. (2006) *Context Aware Ubiquitous Learning Environments forPeer-to-Peer Collaborative Learning*, Educ. Technol. Soc., 9/ 2006, 188-201.

[13] Das M.M., Chithralekha T., SivaSathya S. (2010) *Static Context Model for Context Aware E-learning*, International Journal of Engineering Science and Technology, 2/2010, 2337–2346.

[14] Andronico A., Carbonaro A., Casadei G., Colazzo L., Molinari A., Ronchetti M. (2003) *Integrating a Multi-Agent Recommendation System into a Mobile Learning Management System*, Proc. of Artificial Intelligence in Mobile Systems 200, October 12, Seattle, USA.

[15] Rosaci D., Sarné G. (2010) *Efficient Personalization of E-learning Activities Using a Multi-Device Decentralized Recommender System*, Comput. Intell., 26/2010, 121-141.

[16] Zaïane O.R. (2002) *Building a Recommender Agent for E-learning Systems*, *Proc. of the 7th Int. Conf. on Computers in Education*, Auckland, New Zeland, 55-59.

[17] Zakrzewska D. (2011) *Building Context-Aware Group Recommendations in E-learning Systems*, LNAI 6922, Jędrzejowicz P., Nguyen N.T., Hoang K. (eds.): Computational Collective Intelligence - Technologies and Applications. ICCCI 2011, Part I, pp. 132-141.

[18] Dey A.K. (2001) *Understanding and Using Context*, Pers. Ubiquit. Comput., 5/2004, 4-7.

[19] Das M.M., Chithralekha T., SivaSathya S. (2010) *Static Context Model for Context Aware E-learning*, International Journal of Engineering Science and Technology, 2/2010, 2337–2346.

[20] Viola S.R., Graf S., Kinshuk, Leo T. (2007) *Investigating Relationships within the Index of Learning Styles: a Data Driven Approach*, Interactive Technology & Smart Education, 4/2007, 7–18.

[21] Felder R.M., Silverman L.K. (1988) *Learning and Teaching Styles in Engineering Education*, Eng. Educ., 78/1988, 674–681.

[22] Index of Learning Style Questionnaire, http://www.engr.ncsu.edu/learningstyles/ilsweb.html

[23] De Marsico M., Levialdi S. (2004) *Evaluating web sites: exploiting user's expectations*, Intern. Journal of Human-Computer Studies, 60/2004, 381-416.

[24] Zakrzewska D., Wojciechowski A. (2008) *Identifying students usability needs in collaborative learning environments*, Proc. of 2008 Conference on Human System Interaction, Cracov, 862-867.

[25] Zakrzewska D. (2008) *Validation of Cluster Analysis Techniques for Students' Grouping in Intelligent E-learning Systems*, Proceedings of 14th International Congress of Cybernetics and Systems of WOSC, Wroclaw, Poland, 893-901.

[26] Witten I.H., Frank E. (2005) *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd Edition. Morgan Kaufmann Publishers, San Francisco, CA.