Rule-based model for selecting integration technologies for Smart Cities systems

CEZARY ORŁOWSKI

IBM Centre of Advanced Studies in Campus

Gdańsk, Narutowicza 11/12 Koszalin, Śniadeckich 2, Poland e-mail: cor@zie.pg.gda.pl

INTRODUCTION

The aim of this paper is to present the stages of developing an information technology integration model for the design of systems for Smart Cities. The introduction explains the need for integration technologies and presents a model for selecting integration technologies and the conditions for its use in designing Smart Cities systems. Then two verification stages of the presented model are discussed. In the first stage, the model underwent a replicative verification in the Eureka project, the purpose of which was to build fuzzy decision-making scenarios for the needs of Smart Cities. In the second stage, the developed model underwent a predictive verification process which aimed at selecting integration technologies (suggested by the developed model) in the construction of an IOC (Intelligent Operating Centre) decision-making system. The paper is summarized by conclusions and observations about the importance of integration technologies and the possibility of predicting them.

THE DEVELOPING OF SMART CITIES SYSTEMS

The process of developing information systems for Smart Cities is complex, lengthy and involves considerable and normally dispersed resources (Bhowmick et al. 2009). For this reason, the developing of such systems is risky and requires difficult decisions to be made by the management. The analysis of these projects indicates that two processes, which are key processes for their success, must be taken into account, namely management of the developing team and management of information technology (Bass et al. 2009).

In the case of management of the developing team, the problem lies in the long-term duration of the project, the need for cooperation between experts from various fields and the knowledge of information technology which changes in time as well as in use. As far as information technologies are concerned, their correct selection and the need for integration processes becomes evident. While a considerable amount is written on the subject of IT management processes (Czarnecki et al. 2009), (Kowalczuk and Orlowski, 2007), (Bass et al. 2009), there are much fewer publications which refer to the integration of technologies and IT resources in Smart Cities systems.

Smart Cities systems have the potential to fully support the decision-making processes of large urban areas and can integrate a tremendous amount of data coming from different sources. If we also assume that such data will flow into the system in a long-term manner, problems connected with the standard of data, the methods for its processing and especially its integration prove to be critical for ensuring the maintenance processes of these systems. The appropriate selection of integration technologies will provide not only the supply, but also the processing of the incoming data. (Orlowski and Kowalczuk, 2006).

The paper incorporates the author's many years of experience in the field of information technology management, the result of which was the model of information technology management presented in many previous works. This paper includes a detailed model of selecting integration technologies for IT systems. This model is also a generalization of experience in the field of IT project management. It can be used to predict integration technologies for both databases and developing environments for the design of Smart Cities systems and their architectures.

IT TECHNOLOGIES INTEGRATION MODEL

The presentation of the model will start with its assumptions. The model is designed for project managers who are forced to make decisions regarding the integration of IT resources during the course of a project. The concept of resources refers to database technologies and developing technologies. Database technologies (from the view point of the client's representatives, the supplier and the experts cooperating with the project team) have a decisive influence on the way data is placed and processed for the construction of decision-making scenarios. The more spacious the database technologies are, the richer the scope of their use and of the decision-making scenarios. This is why it is so easy to implement the *. SQL or *. DB2 database standards.

Data entered into the system is then processed. If a uniform database standard is provided (which is the case for simple systems), then both the input and processing of data do not constitute a problem from the point of view of system development. If, as is the case in large systems which process huge resources, varied data is entered, then the method of database integration, and consequently the integration of developing technologies, decides on the development of the system. These processes can be easily represented as a feedback system in which integration technologies tune this system in terms of the efficiency of the decision-making. The diagram of the feedback system is shown in Figure 1.

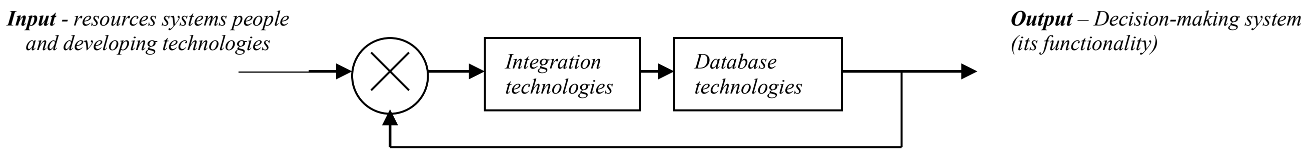

Fig. 1 The integration technologies selection process as a feedback system

The input to the system is data developing team and developing technologies used by this team. Quite a lot can be said about the team by taking into account its maturity (in the sense of CMMI - *Capability Maturity Model Integration*), (Pastuszak et al. 2008). The author's own work in this area (Orlowski et al. 2009), (Orlowski and Kowalczuk, 2006), (Orlowski and Szczerbicki, 2002) also points to the importance of the maturity parameter in the evaluation of developing teams. It is assumed then that the maturity parameter aggregates the experience, the knowledge and the way in which the developing and management processes are implemented by the team.

The second component of input to the system are the developing technologies used by it. It should be assumed that they derive from maturity; however, there is no direct connection between the maturity of the team and the type of technology used. Mature teams can use less technologically advanced solutions (IT tools or programming languages), but can also apply very sophisticated CASE tools (Computer Aided Software Engineering). Similar trends can also be observed in the case of teams with low maturity. Unfortunately, such an approach to the use of information technology is then reflected in a simple relationship. The information technology applied is one that the team specializes in and it is changed very carefully. The range of available technologies, however, allows the choice of both types (languages, tools, developing environments) as well as specific technologies (Java, .Net, Eclipse, J2EE, etc.).

The IT technologies for developing used by the developing team are reflected in the way database technologies are used. While in the case of simple systems the supplier imposes the database standard, for heterogeneous systems (with different database standards) imposing a uniform standard can significantly reduce both the design process as well as the development of the system. Therefore, warehouses are increasingly being used as a method of data integration whereas properly designed ETL processes (*extraction, transformation, loading*) allow the input and processing of any data in the system. The question is: to what extent will data warehouses be able to allow the introduction of any database standards typical in the design of Smart Cities systems? For this reason, there are more and more integration technologies based on ESB (Enterprise Service Bus), which is the answer to the above question. Therefore, the applicability of the data buses, warehouses or a simple database integration depends mainly on the experience of the project manager.

Similar questions arise when determining the functionalities of the decision-making system. Are they limited only to monitoring events or is their scope partial, covering both the defining of events and incidents (event groups) as well as the identification of key performance indicators? It can also be assumed that the range of functionalities is full, which means that, in addition to those mentioned above, the system should also generate notifications, alerts and event rules thanks to the construction of standard operating procedures (SOP), as well as carrying out analyses on the level of *Business Intelligence*.

The answer to questions like these lies in the proper selection and application of integration technologies. Their scope, which was briefly mentioned above, needs to be extended by the SOA (Service Oriented Architecture) approach to designing service-based architectures. If we take into account designing the SOA together with the service bus, data warehouses, or a simple database integration, then (including product information) we have a compendium of knowledge about integration technologies. What is left then is only the choice of the appropriate integration technologies for the developing processes. In such a case, it is possible to generalize the system shown in Figure 1 to a model of the overall selection of integration technologies for ITMI (IT Model of Integration) project teams. Such a general model is shown in Figure 2. To describe the scope and the value of each variable, a linguistic description can be used preceded by an expert evaluation as to what maturity / height the technologies belong.

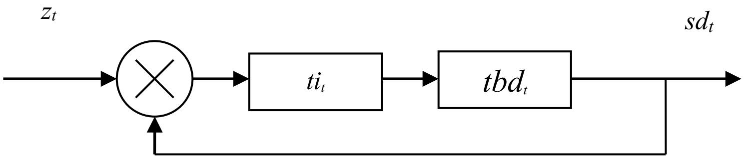

Fig. 2. A general model of ITMI integration technologies selection in the developing processes of Smart Cities systems

A scalar description was used for this model (to be simpler) and, as before (Figure 1), the general model was used as the control system with the function f_t assigned to it.

$$sd_t = f_t(ti_t, tbd_t) \tag{1}$$

where:

f_t – control function (rule-based function proposed in the paper)

sd_t – output object response (utility level of the decision-making system), $sd_t \in \langle 1, 5 \rangle$

ti_t – the maturity of the controller (integration technologies), $ti_t \in \langle 1, 5 \rangle$

tbd_t – the maturity of the controlled object (database technologies), $tbd_t \in \langle 1, 5 \rangle$

z_t – set value (the maturity of the team and the technologies of developing), $z_t \in \langle 1, 5 \rangle$, t - time

While working on the generalization of the model, a vector description can be applied which takes into account the existing compendium of knowledge about the developed model. Then the vector of the maturity of the database technologies can be expressed by the following formula (2)

$$\mathbf{tbd}_t = \begin{bmatrix} mtbd_t \\ ntbd_t \end{bmatrix} \tag{2}$$

where:

$\mathbf{tbd}_t$ - the vector of the maturity of the database technologies

$mtbd_t$ - the variable of the method of the database technologies, $mtbd_t \in <1, 5>$

$ntbd_t$ - the variable of the database technology tools, $ntbd_t \in <1, 5>$

On the basis of the gathered knowledge, a more complex vector of integration technologies can also be determined, which is so important from the point of view of constructing the ITMI model.

$$\mathbf{tti}_t = \begin{bmatrix} tbaz_t \\ thur_t \\ thbus_t \\ tbus_t \end{bmatrix} \tag{3}$$

where:

$\mathbf{tti}_t$ - the vector of the integration technologies

$tbaz_t$ - the variable of data integration via databases, $tbaz_t \in <1, 5>$

$thur_t$ - the variable of data integration via database warehouses, $thur_t \in <1, 5>$

$thbaz_t$ - the variable of data integration via database warehouses and a service bus, $thbaz_t \in <1, 5>$

$tbus_t$ - the variable of data integration via a service bus, $tbus_t \in <1, 5>$

The full scalar and vector description (partially presented in the paper) creates conditions for the implementation of the ITMI model (Nguyen and Szczerbicki, 2010). The f_t function can (for both the scalar and vector description) be represented in a rule-based form. In the case of a scalar description, the rule-based description will contain two input variables and one output variable. Assuming the linguistic description used in this paper and the later fuzzy description (considered to be applied), the complete RITMI *(Rule IT Model of Integration)* model of IT technology integration will contain 2^5=32 rules (Zadeh, 1978). An example of the implementation of the model with the use of the rule-based description is shown in the equation (4)

IF the integration level < ti_t > is at <1> and the technology level as < tbd_t > is at <1>, then the usability of the DSS system < sd_t > is at <1> (4)

The applicability of both scalar and vector descriptions also means that the general RITMI model (Fig. 2) can (due to the compendium of knowledge about the state of the integration technologies) be used for the selection of these technologies in the developing processes of Smart Cities systems. In this case, we have the knowledge of types of given technologies. Moreover, if we are able to present the structure of the model (Fig. 1 and 2), then the RITMI model can have replicative and predictive validity (Czarnecki and Orlowski, 2009). Therefore, the following step in the presentation of the model is to demonstrate those validities.

EXAMINATION OF THE REPLICATIVE VALIDITY OF THE MODEL

The developed RITMI model underwent the processes of replicative verification in the environment of the IT project Eureka *Air quality management in urban areas using a Web Server E!3266*. Fuzzy decision-making scenarios were created in this project, which were then implemented in the Webair_PM decision-making system developed for various Smart Cities projects. When preparing data for the project, a precise description including both data and the database system was developed. Two different solutions were taken into account in terms of database construction strategies. In the first solution, a direct feed of data from the project partners to the Webair_PM system was assumed. The second solution was based on building a data warehouse powered from external databases of the project partners. Before choosing a solution, both versions were tested.

On the one hand, the first situation required an analysis of the different database standards and on the other, the possibility of feeding the constructed system with data of varied standards. Therefore, two experiments were carried out. In the first one, data requirements and standards were specified (Armaag - pollution data, the City Council - noise data, Gdansk University of Technology - pollution and weather conditions.) A parallel database feed via Webair_PM was also carried out.

It soon became apparent that the parallel feed was very difficult to realize, and therefore it was decided to perform the initial feed with batch data (experiment II) to assess the suitability of the data and the mechanisms of its acquisition. In the case of batch processing, there was no problem with obtaining the data, but there was a problem with reproducing the batch feed. Therefore, a more complex solution was found which involved building an external entity in relation to the Webair_PM system, namely a data warehouse. Processes for supplying both the warehouse and Webair_PM were developed. Although the mechanism proved to be useful for Webair_PM, it was not possible (for data from Armaag) to provide a continual new data feed to the warehouse.

The standard of the data warehouse was also analyzed (from the perspective of the validity of the RITMI model). Two file standards, *.SQL and *.DB2, were taken into account. From the point of view of Webair_PM, the better solution (meeting the requirements of Webair_PM) was the former standard. However, the experience of members of the team was taken into account, to whom the DB2 standard was unfamiliar. Therefore, the SQL standard was finally chosen, although DB2 seemed to be more developmental. The decision to choose a more familiar standard over a more predictive one stemmed from the need to reduce project risk. The diagram of the data integration processes (according to the suggestions of the RITMI model) of the Eureka project is shown in Figure 3.

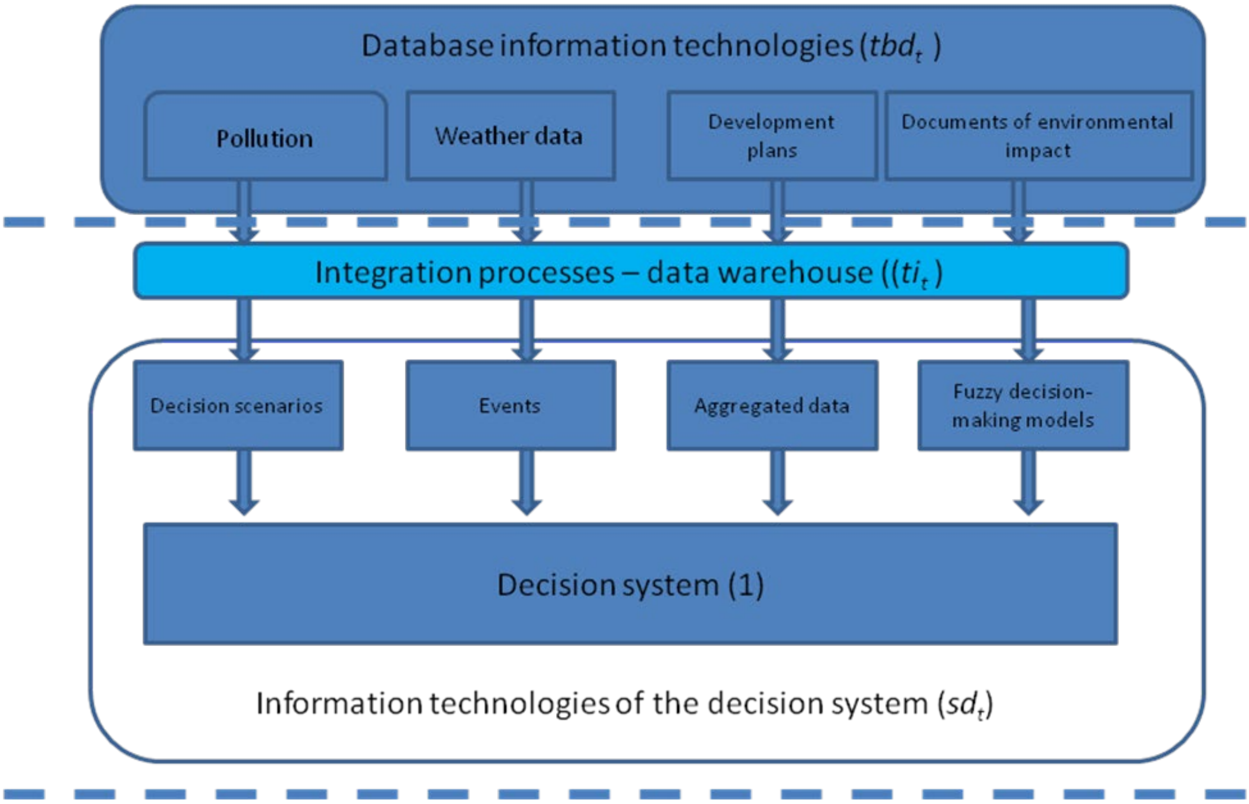

Fig. 3 Data warehouse, as the integration technology in the Eureka project (selected according to RITMI).

The study of replicative validity involved identifying both the components of the model (the database system and integration technologies), the relationship between these components as well as demonstrating that integration technologies control the process of constructing a decision-making system. While the identification of the two components proved to be relatively simple (the resource of the SQL standard database, the data integration processes and data warehousing), the relationship between these two components proved to be a much more complex problem. To support it, it was necessary to create a database trigger for the data warehouse. It turned out that its use facilitated the flow and the feed through the ETL processes.

In the process of replicative verification, it proved equally difficult to demonstrate that the use of a data warehouse is a controller of change in the processes of constructing the Webair_PM decision-making system. The transition between the different integration technologies observed during the construction was a process resulting from the experience of the managers of the project rather than a sequential process of formal changes. Therefore, the replicative verification process should be considered partially successful.

EXAMINATION OF THE PREDICTIVE VALIDITY OF THE MODEL

The predictive verification process was carried out in the project of constructing the IBM IOC (Intelligent Operating Centre) decision-making system (Smith, 2013). The IOC is an intelligent city management system which (because of its functionalities) uses data collected from multiple sources (from environmental monitoring, industrial network resources, crisis management repository (cameras, security systems, early warning systems and others)).

After an analysis of the requirements regarding pollution (for Gdansk), the concept of the project and the construction of the decision support system, and on the basis of the RITMI model suggestions, (Fig. 2) the requirements for the IOC integration technologies were determined. These technologies had to be applied both at the level of individual servers as well as at the level of the existing data warehouses. Thus, a decision was made (based on the suggestions of the RITMI model) about the selection of the integration technologies - the ESB service bus and, consequently, the selection of the IOC system architecture as one based on the SOA. Data architecture was designed on the basis of the *.SQL files and the architecture of the applications was based on RAS (*Rational Software Architect*). The integration service bus was built on the basis of the *WebSphere Message Broker Toolkit* (bus structure) (Fig. 4). Then the resources of web servers were made available (City Council, Armaag Foundation, Gdansk University of Technology). In the case of

Armaag, permanent access to the bus could not be successfully established (the ESB was supplied with batch data).

In the case of data from Gdansk University of Technology, a similar solution was used as the one for the City Council. The server and its resources were identified and then the service bus for the IOC was fed on-line. In addition, integration tests of noise level maps were carried out (for the existing data). It was assumed that the process of obtaining the maps for the ESB bus would have two stages: in the first, maps would be delivered in the form of batch files, in the second, they would be fed to the system via access mechanisms built in with the help of the *Message Broker Toolkit*.

The ESB validity verification process also included the possibility of using decision-making scenarios, the correct operation of the warehouse (connected to the bus) and two functionalities available from the level of the ESB bus, namely: the presentation of events (noise and its levels, pollution and its levels) and the presentation of maps (the noise map) applied onto the map of Gdansk (Fig. 4).

Apart from testing data and scenarios which can be analyzed, the possibility of integrating the IOC environment to the ESS was tested, while exploring the possibility of constructing scenarios on the basis of internal testing mechanisms (attached to the bus). The process involved three stages:

- Searching for implementation environments of inference mechanism scenarios
- Searching for IOC support mechanisms in constructing scenarios
- Possibilities of building own inference environments for the scenarios and data from IOC

The analysis of the results shows that all three approaches can be implemented, which proves that the choice of the bus as the project resource integration environment was the correct one. Figure 4 presents the IOC system construction environment selected on the basis of the RITMI model indications.

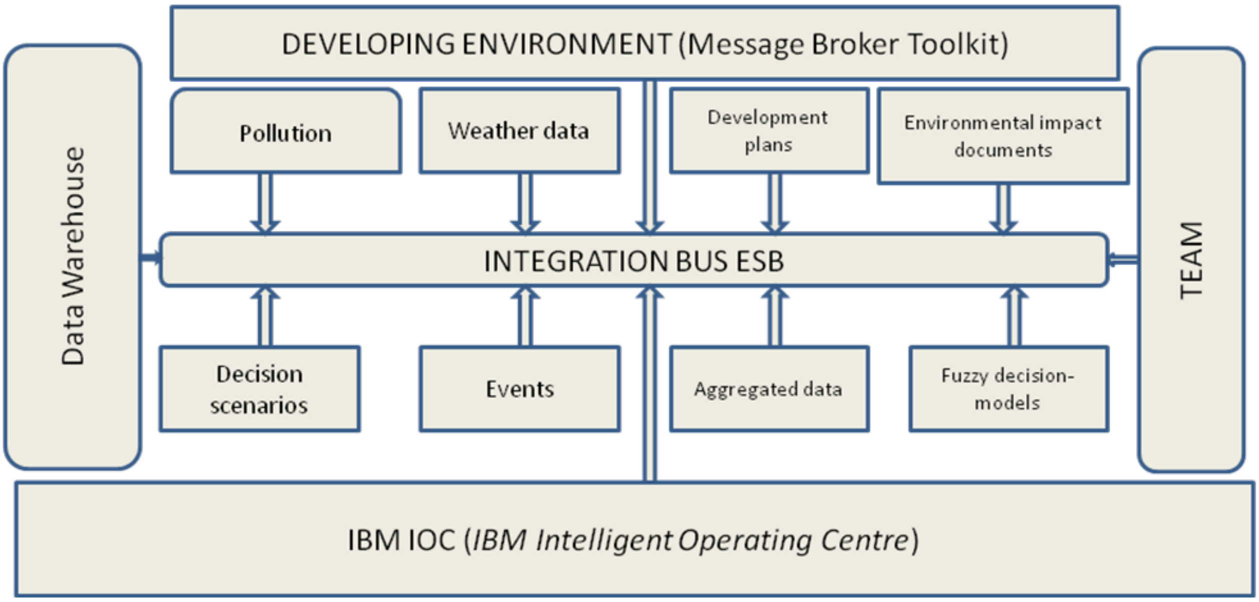

Fig. 4 The integration bus, as the integration technology in the IOC project (selected on the basis of RITMI).

CONCLUSIONS

The aim of this paper was to present the stages of the construction of the RITMI model for Smart Cities systems. The choice of integration technologies is crucial for the development of these systems due to the long time of their construction, the huge resources and the different database standards. The selection is also important from the point of view of the cooperation of experts from many different fields, such as planning, pollution and risk analysis.

Therefore, the initial part of this paper presents the stages of constructing the model, pointing to the knowledge resources needed for its construction, but also for its verification. The evaluation of its validity for IT project managers lies in the possibilities of its verification. In their work, many IT project managers rely mainly on their experience, assuming that there are no models or systems which could help them. Therefore, the aim of this paper was to show that combining knowledge and experience can lead to the creation of such models and their subsequent use in practice.

The replicative verification showed that the practice of project management takes into account the dynamics of the changes in integration technologies, although it is not able to identify the direction of these changes and, in particular, the possibility of implementing these changes. The environment of the Eureka project described in the paper showed, in fact, that there is a need for applying a warehouse as the next stage of resource integration. This environment, however, did not take into account the need for a precise definition of the ETL processes, as was demonstrated in the process of verification.

The predictive verification carried out in the IOC project clearly indicated the purposefulness of using the resource integration model. However, it is worth noting that although the integration service bus was indicated, not all resources could be connected to the bus. For a project manager (a person using this model) information on the selection of the ESB and the construction of the architecture of the IOC decision-making system on the basis of SOA is sufficient to carry out the project work. Therefore, in this scope, the applicability of this model is significant.

The expansion of the model is planned mainly in the field of linguistic description and assigning linguistic values to the variables of the model in a greater extent than was presented in the paper. This will be possible when project managers start to predict the application of integration technologies on the basis of the RITMI model in their work.

REFERENCES

Bass, L. Clements, P. Kazman, R. 2003. *Software Architecture in Practice*. Boston: Addison-Wesley.

Bhowmick A. 2009. *IBM Intelligent Operations Center for Smarter Cities Administration Guide* 5. New York: International Business Machines Corporation.

Brandon, L. Shrum, S. 2009. *CMMI for Services: Guidelines for Superior Service*. Boston: Addison-Wesley.

Czarnecki A., Orłowski C., Sitek T., Ziółkowski A. 2009. Information Technology Assessment Using a Functional Prototype of the Agent Based System. *Foundations of Control and Management Sciences,* 23:7-28.

Czarnecki A., Orłowski C. 2009. Ontology as a Tool for the IT Management Standards Support Agent and Multi-Agent Systems: *Technologies and Applications*: 330-339.

Kowalczuk Z., Orłowski C. 2007. Knowledge Management Based on Formal Models of Software Project Management. *Cybernetics and Systems,* 38: 611-626.

Nguyen N., Szczerbicki E. 2010. *Intelligent Systems for Knowledge Management*. Berlin Heidelberg: Springer-Verlag.

Orłowski C., Kowalczuk Z. 2006. Knowledge Management Based on Dynamic and Self-Adjusting Fuzzy Models*: Knowledge-Based Intelligent Information and Engineering Systems*. Springer Berlin Heidelberg.

Orłowski C., Ziółkowski A., Czarnecki A. 2011. Validation of an Agent and Ontology-Based Information Technology Assessment System. *Cybernetics and Systems*: 41: 62-74.

Orłowski C., Kowalczuk Z. 2006. Knowledge Management Based on Dynamic and Self-Adjusting Fuzzy Models. *Lecture Notes in Artificial Intelligence,* 68: 671 - 678.

Orłowski C., Szczerbicki E. 2002. Evaluation of Information Technology Projects. *Cybernetics and Systems,* 33: 659-673.

Pastuszak J., Stolarek M., Orłowski C. 2008. Concept of Generic IT Organization Evolution Model*,* *Faculty of ETI Annals,* 18: 235-40.

Smith A., D. 2013. *IBM Intelligent Operations Center KPI Implementers Guide for Websphere Software*. New York: International Business Machines Corporation.

Zadeh L. 1978. Fuzzy Sets as a Basis for Theory of Possibility. *Fuzzy Sets and Systems* 1: 103-111.