Within the problem of making the decision based on the document management system

Alexander Kurbatski
Vice-rector of the Belarusian state university
4, Scoryny av., Minsk, Belarus, 220050,
phone: (375-17) 2206680, fax (375-17) 2265940, e-mail: kurb@unibel.by

ABSTRACT

This report is dedicated to the one of the central problems of modern communications – the problem of document processing automation and principles of the creation of the decision-making systems based on this. This problem affects all levels of human and society intellectual activities. It's solution is required by all hierarchic levels of functioning of the social systems and groups.

Modern workflow systems and technologies are characterized by the major functional and corporate principle – corporativeness. This principle stipulates the changing of the workflow automation approaches (in comparison with the previously known approaches).

Architecture of any of the modern corporate information system includes as a base business operations automated system as well as system of the management of the documents in electronic form.

To solve this problem methods and techniques developed in the field of computer science are widely used.

The architecture of any modern corporate information system as a basis includes

Computer Science appeared not so long ago, significantly later than others sciences. During the first stage of its development it "learned from it's neighbors". It used ideas and methods of related as well as distant areas to solve its problems. However, information technology were high developed recently. Now it is more complex and "scientific" than its "neighbors". It became sharper and closer to optimal. In this field computer science is a leader among relevant sciences. And now the "opozite direction movement" is starting: methods used in computer science are transported to the others. One may claim that both principles: "Do it with the assistance of the Computer" as well as more complex "Do it as well as computer does" are right.

It is proposed that some methods used in computer science are to be used in optimization of the technology in other relevant fields.

These methods are vectorizing, pipelining, and multiplexing. They are quite simple and already used by engineers and technologists in development of the technological conveyers. So the results described here are not brand new. The goals are to create a "bridge" between discussed directions, stress existing similarities. Further to the computer science development all these could be effectively used in future. This paper has, in major, methodological features. But it still contains some pragmatic aspects.

Firstly, Methods and algorithms of vectorizing and pipelining developed in computer science has essential theoretical base (mathematical theorems). Using them leads to the confidentiality that everything that received from the output is correct and optimal. Whereas no one could be confident when using nonsystematic methods of synthesis and optimization appeared in each case.

Secondly, there are a lot of software implementation of the proposed methods. While describing initial data using required formats, the solution of the problem could be received automatically (or semi-automatically using dialogs) and in comparison with the traditional methods much faster. Unfortunately it is not easy to find appropriate software system.

Finally, it is important that there exists a good experience in use of the developed methods not in the field of computer science.

Thus the basis for the solution of the problems of the search of the best for the given criteria or optimal document management schemes and systems is created.

In the widest sense, without economical criteria, the problem of optimization of the document processing using modern information technology is bound to the optimization of the complex systems. At least, the following aspects of the domain allow clarify the class of the optimization problems.

Firstly, the optimization approach could be based on the principles of the mutual reflection of the knowledge domain and the resources used for process automation in the named knowledge domain. At first it is necessary to select the workflow concept, formulate the goals, sub goals and criteria of their achievement, specialize features and parameters of the manipulation process, and formulate the limitations.

Finally, it is necessary to know that only the aggregate of the models for the automation object permits to fully describe object's function as a set of the functional links. This suggests the idea of the decomposition of the automation object and solution of the problem divided into parts. The solution received will be quasioptimal.

The use of any of the described above methods includes the solution of the three major problems:

1. The workflow system analysis
2. Optimization
3. Synthesis(selection) of the optimal workflow system

Let's consider the core of each problem.

Workflow analysis assumes the quantitative and qualitative evaluation of the system. Some analysis stages are considered as a reflection of the (projecting, functioning) system into the formal model class. This class permits to receive the evaluation of the workflow system parameters, for instance, probability evaluations of the queues, system breakdowns, document processing time, overloaded unloaded pipes etc.

Optimization aims at the workflow system or its parts modifying. The main goals of the process of optimization are

a) decreasing of the single document processing time
b) organization of the document flow processing in their arriving order
c) queue liquidation
d) replacement of the non effective links, or in the terms of the managerial workflow the staff movement, deletion of the non-loaded links (staff dismissal)

Optimization is performed in accordance with one or several criteria. One of the actual problems is the problem of the optimal correlation of the system

timing parameters (single document processing time and flow processing time) required resources, for example the number of the links in the executing chain.

The problem of optimization of the workflow system working with the limited resources is typical. For example the problem requires that:

a) there will be no staff dismissal
b) the number of the executor will not be higher than the limitation estableshed
c) the number of the executing links in the system reduces

The required time parameters are achieved through the methods of the document manipulation processes vectorization.

Finally, using the modern capacity of the information technology the workflow system synthesis is the selection of the architecture and system construction using of pluggable modules. The set of the modules and their linkages within the selected architecture permit to implement any document flow processing and managing technology.

The approach based on the decomposing of the objects permits flexible manipulation of the various formalization, modeling and resource allocation methods. Let's assume that the solution of the optimization problem could be performed with the resource limitation or without. The last case is common, when the value of the document produces benefit that is bigger than the expenses spent on the document processing. Although the models used in several projects for the document processing formalization (for example Petri Nets) are by their nature pipelined and conveyered they are the models of the imperative type. They assumes the existence of some external centralized management, which orders processing modes. However, practically it is difficult to fit these requirements. Each processing is performed in accordance with the certain document or group of documents arriving rather than due to the "order from above". The strict order of, even in case of parallel processing could not be set up. Even if it is, it becomes non-effective. The process order is unpredictable due to the unreliable links (illness or unreliability of the executing person), bureaucratic delays, lack of the technical equipment etc. The situation become worth during the period when the document forms, legislation and processing rules are changing.

So, the asynchronous model and its modifications based on the data-flow schemata (or flow schemata) are closer to the reality.

Firstly the models of this type permit to reflect the wider spectrum of the situations of the managerial workflow correctly. In particular, the simple methods could be used to describe the principles of processing initiation in accordance with the document arrival, document complexing and propagation, and paralleling and conveyering of the processing.

Secondly, there exists a certain experience of such model usage. The results of the solution of the system verification, self-purification (documents and their copies are not jammed within the system when they are processed), time optimization including the using systems' redundancy problems are known.

Thirdly, there exists practical experience of the software usage. Such systems components could be used in development of the workflow automation systems.

For example, this asynchronous model is very complex and its description manually using the means of the developing system is very time-consuming, and leads to the big number of mistakes. For easier solution it is recommended to use the means of the described scheme modeling, and the means of the transformation of the traditional consequent algorithmic framework to the flow. This will permit to specify the system, using the common algorithmic language. Moreover this guarantee that the received on the traditional scheme basis the flow scheme will perform the necessary actions.

Besides, let's mention that flow schemes contain methods and ideology of many different models. They are used in parallel-conveyer structures modeling. To solve the problems of queue size estimation and "life time" of fragments (useful functioning) estimation the less complex model could be used – Petri Nets. This model is also describable by flow schemes.

The methods described are used in document management and decision making support system implemented in the Ministry of Education of the Republic of Belarus. In principle, two implementation plans could be defined.

The first: the system is previously optimized using traditional methods and than described as a flow model.

The second: the system is described as a model without preliminary preparation. For these purposes the local rules: "where to get, what to do, and where

to put " are used. They are followed by some conditions.

Both these techniques leads to the equal result.

Within the approach of the automated information systems based on the document management systems creation proposed method of analysis and optimization could be effectively used during the following stages:

¬ Initial processing (selection) and actualization of input and output documented information. On this stage, the problems of optimal arrangement of the data to minimize computational resources, optimize the search etc. are solved as well;

¬ complex processing of the statistical and analytical data to produce new (additional) information;

¬ Use of the informational arrays in the decision making processes, formation of the managing actions in different forms. The analysis of the input data to extract the mostly valuable components and selection of the optimal action strategies based on the prior information are very important problems of this stage.

Developed techniques, algorithms and design development tools based on this method will be used in the following automated information systems:

Informational resources automated systems: in data systematization problems, database logical structure planning, databanks organization etc.

Workflow and document management systems: in different technology components of the document recognition (understanding) technology (including optical recognition and automated text processing), planning problems and task flow management, automated document flow analysis.

Automated information analytical systems: in analysis, systematization, and classification of data from databases and databanks.

Automated systems for decision making: in state modeling of the informatization object, data analysis and elaboration of the new decision-making algorithms.

Preliminary patent-information researches permit to formulate the following conclusions, which prove the necessity of research continuation.

1. Informational approach is one of the most prospective and widely used in solution of the problems of analysis and optimization in different knowledge domains. It has a number of advantages in comparison with other relevant methods and optimization techniques.

The majority of existing methods and algorithms with the given approach has the following essential disadvantage. The information theory apparatus (in probability form) developed for the data transfer channels is used as a theoretical base. Initial data of the real problem is "tuned" to the existing apparatus using the approximations of different kind. So on the very first stage of the problem raising the preconditions of the non-optimal solution are established. Moreover, for each problem class it is necessary to develop the methods of the initial data reflection to the probability informational model.

The proposed method assumes the formulation of the conceptual apparatus, basic definitions and identities for the problems of analysis and optimization. So the models is adequate to the solving problem. For the problems which initial data is in the determined form it is proposed to use functional approach to represent the problem as a data transfer channel.

2. One of the definitive feature of the majority of modern problems (this fully correct for the problems of analysis and management of information resources) is hierarchic nature of the analyzed systems. Decomposition principle plays the important role in analysis and complex optimization of such systems. When it is used the problem of the formal description and registration of informational dependencies and links of the subsystems of different hierarchical level arises. Non solution of this problem leads to the loss of the information and consequentially to the non-optimal solution.

The problem of non-adequateness of the used system models to the real systems is not enough investigated. For example, in document recognition (understanding) systems to describe the link between different hierarchical levels of the natural language the hidden Markov's model is used. However this model is adequate to the language, when the precondition exists that the text is the Markov's sequence. All further research and modifications of the models are limited by this precondition. The analysis of Markov's models using the informational approach demonstrates low effectiveness of the Markov's language models in accordance with the informativeness criteria. The methods of model synthesis are necessary. These methods should take into account a number of parameters of the modeling system including stipulated by their hierarchic nature.

3. Many optimization problems are NP-complete. There is no practically applicable method of their solution when the problems dimension is big. Today different heuristic methods are widely spread. They permit to receive the solutions that are closest to the optimal. The selection of the heuristic depends on the conditions of the concrete problems. That very much complicates the process of their solution. Moreover, the apparatus of association of different optimization techniques for the selection of the best in accordance with the formal criteria is missing.

4. The analysis of the creation and usage of the informational resources processes in educational sphere demonstrates that, in majority of cases, the stages of systematization, data classification based on the theoretically approved criteria, optimal arrangement planning and information management are minimal or even missing.