PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1086 — 2005
Modele i metody zarządzania informacją i wiedzą

Oleg Sosnowski, Vladimir Labotsky, Eleonora Dunjco

The Belarus State Economic University, Minsk

CORPORATE INFORMATION CONTROL SYSTEM OF A UNIVERSITY

1. Introduction

Corporate information systems (CIS) are generally used for planning and management of internal resources of an enterprise which has many different divisions and a network of the affiliated organizations.

The basic purpose, which the enterprises aspire to reach at introduction of modern CIS, is to receive the advantages in the competitive struggle on the commodity market and services.

The activity of a higher educational institution as an object of management is now characterized by a number of factors, the main of which are:

- a variety of sources of financing;
- a variety of types and forms of educational, scientific, industrial and economic activities;
- a variety of forms of information streams;
- a need for the analysis of the market of educational services and labor market;
- territorial dissociation of the divisions and branches.

The specified factors basically define a variety of forms, the character and the volume of information streams of the university that complicates the process of the administrative decisions made by the university authorities.

It is possible to notice that in comparison with the industrial enterprises the purpose of a corporate information university is not so much the improvement of the industrial activity, but the reduction of expenses and efforts for the support of its internal information streams.

Another feature of the university CIS is the necessity of the intellectual information streams management, such as curricula and programs, calculation and optimization of the academic program, the library volume, etc.

Generally CIS are represented by a set of integrated appendices, allowing to create the integrated information environment for automation of planning, account, control and analysis of all basic types of activities of higher educational institutions.

From the point of view of information technologies the university CIS consists of three basic systems: control systems of knowledge – KM-system (Knowledge Management System), control systems of resources – ERP-system (Enterprise Resource Planning System) and systems of support of decision-making – DSS-system (Decision Support System) without which the life cycle of the information would not be completed (fig. 1).

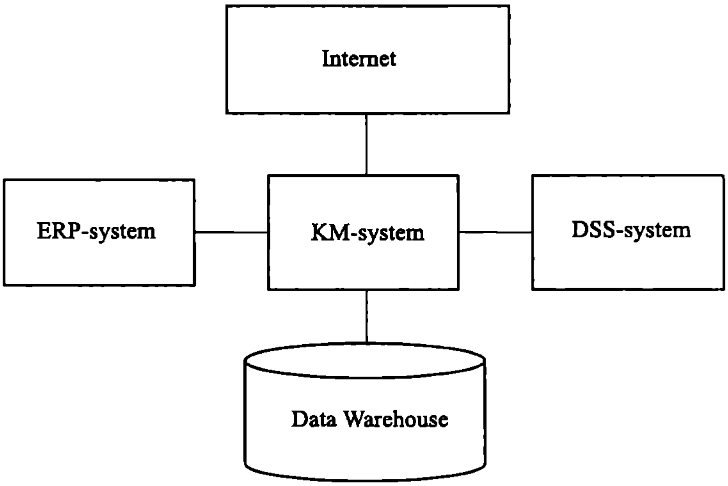

Fig. 1. The University Corporate Information System

2. Data Warehouse

The lowermost level of information system is the storehouse of the data (DW – Data Warehouse) which contains all intellectual property of the university. These may be documents, directories, structural tables, business rules, the description of educational processes, etc.

The quality of the data submitted in the storehouse also depends on the presence of the unified thesaurus – a set of morphological units used for the description of educational process.

Thus, having created the storehouse of the data, we receive the decision of one of the problems of the decision-making process automation: the system where the data are full and submitted correctly. The data shipped in DW, being organized into the integrated complete structure, possessing natural internal communications, get new properties providing them with a status of **information**.

There should not be a direct access to the storehouse both for users and for different university systems. The direct access has only the KM-system which serves as a sluice for other systems and forms an information environment of the university.

The development of DW is a strategic project. Therefore, before creation of the storehouse the number of the decision making people and their information needs are to be studied and analyzed, and after that the requirements for the information storehouse as well as the quantity, quality, structure of the data stored in it introduced from the internal operative systems and the external sources are to be arranged.

With introduction of DW it is possible to go to the next "intellectual" stage – the "extraction" of knowledge. To transform the information and data stored in DW into the useful knowledge the special methods OLAP (*On-Line Analytical Processing*) and DM (*Data Mining*) are used.

3. KM-system

The KM-system is the main system which solves problems of creating in the university a unified information space, provides the organization of team work for the users for acquisition, representation and exchange of knowledge, allowing the access to unified corporate base of knowledge and providing the conditions for an efficient use of their knowledge for common interests.

The KM-system solves five main tasks:

- workflow;
- content management;
- analysis;
- gathering, search and delivery;
- tracking and document circulation.

Workflow. Ways of interaction are necessary for realization of teamwork, for example: Internet networks, technologies of group processing, synchronous and asynchronous conferences, etc.

Content management. Technologies of content management allow users to fix, systematize and organize the ideas in the centralized storehouses simplifying the access to the information.

Analysis. Users should get access to the means necessary for the independent analysis of the corporate data. New concepts of construction and realization of the information systems focused on the analysis of the data, for example: OLAP and DM.

Gathering, search and delivery. One of the problems of the KM-system – association of the groups scattered on different branches of the university or working in the removed mode.

Tracking and document circulation. Control systems of document circulation should provide storage, archiving, indexing, labeling and publishing of the documents.

4. ERP-system

One of the most general problems of the university ERP-system is the problem of distribution and rational use of resource management of the university. Under the university resources one understands the intellectual, financial, material, technical (the number of workstations, liaison channels, the number of servicing devices), time and other resources. Under the types of activity one should consider the educational, research, administrative, industrial, economic, advertising and other types of activities. Distribution of the resources according to various types of activities of the university should provide maximization of "benefit" from the implementation of the results of operation under the conditions of the resource limits and optimized intensity of the types of activity. Besides, the restrictions for the specific resource limits spent for the implementation of a unit of each type of the activities may be imposed.

The well-known classical methods of the mathematical analysis for the solution of the problems for the resources distribution, based on the research of derivatives for the criterion of a function, frequently appear unsuitable for the presence of strong restrictions on variables and the area of change of the criterion function. Methods of mathematical programming provide to overcome these complications [1]. The most practical application was found for the models of linear programming in their classical statement which are more simple for program realization in the automated control systems from the point of view of the organization and databank management.

Formally the problem statement for the distributive type as a model for linear programming could be formulated as follows: to maximize the criterion function

$$Z = \sum_{i=1}^{n} C_i x_i \to \max$$

at restrictions:

$$\sum_{i=1}^{n} a_{ij} x_i \le b_j, \quad i = \overline{1,n};\ j = \overline{1,m},$$

where n is the number of types of activities with the intensity $x_1, x_2, ..., x_n \ge 0$, m is the number of types of resource with the possible limits of consumption amounts by the values $b_1, b_2, ..., b_m$, a_{ij} – the charge of the j resource on i type of activity. The structure of the criterion function $\sum_{i=1}^{n} C_i x_i$ reflects the contribution of each type of the university's activity to the general result.

To maintain a fully adequate problem solution close to the real conditions some elements of the model will be considered as random variables. Thus, the problem may be formulated as the following stochastic model:

$$Z=\sum_{i=1}^{n} C_i x_i \rightarrow \max$$

at probable restrictions:

$$p\left\{\sum_{i=1}^{n} a_{ij}\, x_i \le b_j\right\} \ge \alpha_j,$$

where $i=\overline{1,n}; j=\overline{1,m}, \forall x_i \ge 0; 0 \le a_j < 1;$ α_j – the set level of probability [2].

The problem in probable statement after some transformations may be reduced to a standard problem of linear programming [3]. As a more general result it is possible to consider an integrated parameter of quality training, the university income, the university rating potential, an integrated parameter of the performance productivity of different types of activities considered in the model, etc.

As it has been marked above, it is something common for stating similar problems for some production processes to consider the intensity of the resource use the quantitative definition of which is complicated with the conditions of uncertainty. The opportunity of such estimation is represented by the use of an integrated quality estimation of the particular types of activities. The efforts for the achievement of a particular degree of quality of each type of activity will be accompanied by the charge of the limited resources. Thus, during the current period the presence of the particular resources will be a random variable.

It is necessary to take into account one more feature of the university. As it organizationally represents a system with hierarchical structure, it is very difficult to state the problem due to the need of consideration of splitting the types of activity according to the divisions (departments, faculties), i.e. to provide decomposition.

Thus, it is required to define the distribution of the resources, at which the criterion function

$$Q=\sum_{i=1}^{n} K_i Q_i,$$

(where K_i – weight coefficient which takes into account the importance of a particular type of activity for a particular university during the current moment; $Q=\sum_{l=1}^{k} C_{il}\, q_{il}$ – the integrated estimation of quality i type of activity of all divisions) will accept the maximum value and restrictions will be executed:

$$p\left\{\sum_{i=1}^{n} a_{ijl}\, q_{il} \le b_{jl}\right\} \ge \alpha_{jl} \sum_{l=1}^{k} b_{jl} \le b_j \quad l=\overline{1,k};\ j=\overline{1,m},$$

where k is the number of the university divisions, C_{il} – the specific cost of improvement of quality i type of activity in *that* division; q_{il} – the quality level i type of activity in l division; b_{jl} – the size of j resource, provided by the university for the improvement of the activity quality l of the divisions; α_{jl} – the set level of probability of the j resource to maintain the qualitative activity of l divisions.

The offered model provides the solution of problems of time balanced distribution of the resources according to the type of activity both on the university and its divisions levels. The model may also be used for the definition of the need for the introduction of new types of activities (services). Managing on the behalf of the decision making person in this model are not the separate parameters as, for example, as the volumes of the resources involved for different needs, but the formation of the procedures for the assessment of the university and its divisions activity. The formulated general problem of the resources distribution and the offered stochastic model demand further development and gathering of the necessary initial information.

5. The intellectual analysis of the data

Data mining is a new technology of the intellectual data analysis for the purpose of revealing the latent regularities as significant features, correlations, tendencies and patterns [4]. There are six standard types of regularities:

- classification;
- clustering;
- association;
- sequence;
- forecasting;
- deviation detection.

Data mining use a the big number of various methods or their various combinations. We list the most important and frequently used methods:

- correlation analysis;
- regression analysis;
- ANOVA/MANOVA;
- cluster analysis;
- discriminant analysis;
- factor analysis;
- decision trees;
- neural networks;
- case-based reasoning;
- algorithms limited searching;
- logic regression;

- fuzzy logic;
- genetic algorithms;
- evolutionary programming;
- visualization.

6. The assessment of the model utility

The automated reliability of the received knowledge assessment, the estimation of the statistical importance of the constructed models is a necessary and very important point of any research. The assessment of the importance of certain dependences of the models may be carried out with the help of standard statistical methods.

On the research of the statistical reliability of the result the values of a standard deviation and a standard mistake are considered. If the value of a dependent variable for i record is $p_i \ (1 \le i \le N)$, and the value of same variable predicted by the found model is P_i the standard deviation is defined as

$$S_{\text{dev}} = \sqrt{\frac{\sum_{i=1}^{N}(p_i - P_i)}{N-1}}.$$

Here N is the number of records for which the regression model is counted. The standard mistake as predicted by the given model variable is defined as

$$S_{\text{err}} = \sqrt{\frac{\sum_{i=1}^{N}(p_i - P_i)}{(N-1)\sigma}},$$

where σ – a square of a dispersion of values p_i:

$$\sigma = \frac{\sum_{i=1}^{N}(p_i - \bar{p})}{N-1},$$

where $\bar{p}$ is the average value of this variable. In other words, the standard mistake is the standard deviation divided by the dispersion.

The most significant, or the most exact model found by data mining methods is the model possessing the least value of a standard deviation among all found models.

The value is a measure of probability that the dependence found "is true" and indeed characterizes the data under the research. For that the negative logarithm of probability is used showing that the dependence is received only by chance as the result of the statistical fluctuation in the data. To calculate the index of the importance the standard deviation of the result received on the real data is compared with a standard deviation of the result received for the data created artificially where the value of a target variable for different records with a casual image is hashed. If the standard deviation received on the real data S_{real} is approximately equal to a standard deviation of the casual data S_{rand} the index of the importance is close to unit, that is the result of the research cannot be considered as significant. If S_{real} is much smaller, than S_{rand} for all casually generated tables, the index of validity is much more than 1.0. In this case the result of research can be named significant. In practice it is meaningful to name significant only those models the index of validity of which is more than 2.0.

On the assessment of the statistical importance of a model some measure of its accuracy is calculated, but one should note, that certain value of the importance is applicable only to those data on which the model is constructed, and to be ready that the data for which the model will be further applied, can differ from initial in the unpredictable image.

References

[1] Мильнер Б.З. *Управление знаниями.* М.: ИНФРА, 2003.

[2] Васильев В.Н. *Модели управления вузом на основе информационных технологий.* Петрозаводск: Изд-во ПетрГУ, 2000.

[3] Моисеев Н.Н. *Математические задачи системного анализа.* М.: Наука, 1981.

[4] Дюк В., Самойленко А. *«Data Mining»: учебный курс.* СПб: Питер, 2001.

KORPORACYJNY SYSTEM INFORMACYJNY UCZELNI WYŻSZEJ

Streszczenie

Korporacyjny system informacyjny uniwersytetu może być stosowany do planowania i zarządzania wewnętrznymi zasobami przedsiębiorstwa składającego się z wielu działów i funkcjonującego w sieci powiązanych organizacji. Działanie instytucji należącej do sektora wyższej edukacji jako obiektu zarządzania można scharakteryzować przez grupę czynników krótko opisanych w artykule. Głównym celem artykułu jest analiza koncepcji korporacyjnego systemu informacyjnego działającego w ramach uniwersytetu jako połączenia trzech podstawowych systemów: systemu zarządzania wiedzą, systemu klasy ERP oraz systemu wspomagania decyzji. Ocenę takiego modelowego ujęcia zaprezentowano w zakończeniu.