Journal of Universal Computer Science, vol. 11, no. 2 (2005), 271-284
submitted: 31/7/04, accepted: 15/12/04, appeared: 28/2/05

Creation of Information Profiles in Distributed Databases as a Game Problem

Juliusz L. Kulikowski
(Institute of Biocybernetics and Biomedical Engineering PAS
4, Ks. Trojdena Str., 02-109 Warsaw, Poland
jlkulik@ibib.waw.pl)

Abstract: There is considered a problem of information profiles and information resources collection forming in distributed data- and/or knowledge bases as a result of an attempt to satisfy the information requirements of the customers represented by their information profiles. It is shown that the interests of managers of data- and knowledge bases are not fully convergent and that they participate in a composite, partially co-operative, partially non-co-operative *n*-persons games. There is given a formal description of the strategies used in such games, as well as the methods of decision making of the players on the level of open access (*OADB*) as well as on local (*LDB*) databases one.

Keywords: distributed databases, information resources, information market, game theory, strategy of database managers
Categories: H.2.7, H.3.2, H.3.5, I.6.8, K.6.4

1 Introduction

A spontaneous and dramatic development of open-access data- and knowledge bases in the last decades has led to a new situation in many areas of human activity depending on the access to information resources. This remark concerns various branches of scientific research, technology, education, administration, trade, health services, national security, natural environment protection, etc. For the existence and development of all the above mentioned areas access to information resources satisfying specific requirements of the users is necessary. Distributed databases accessible through the Internet (or through any other computer networks) make it possible to reach higher quality of human activities and, as a result, they open a new era in development of modern civilisation. But at the same time their impact on our life has caused new problems that never have existed before or existed only in a germinal state. Till the systems of databases were created by single organisations they were dedicated to the interests of database users represented by database managers; the area of possible conflicts between them was then strongly limited. New situation arose when government or other higher administrative authorities tried to initialize design and construction of computer-based information systems in order to force higher effectiveness in sub-ordered organizations. The goals of information systems' sponsors, designers and users became divergent and, as a result, many so-designed information systems failed as being not accepted by their potential users. Such a situation arose, for example, in some East-European countries, where for many years a tendency to built big computer-aided management systems within the central national development programs was dominating. Despite the fact that central planning gives a (theoretical) possibility of optimisation of development programs, in fact, the so-realised information systems are rather far from optimum, because no common understanding of the "optimum" by system's sponsors, designers, managers and users could be established. Free information market gives another possibility of information systems creation according to the requirements of various subjects. In this case existence of no general "optimum" is assumed; information creators, systems managers, and information users can express their own goals and they participate in a *n*-person game trying to reach their individual optima. Simultaneous reaching of all so-defined optima is impossible; however, the market mechanisms make it possible to reach a common balance-point being a compromise between the partners' expectations.

We owe the general concepts of the theory of games to J. von Neumann and O. Morgenstern [Neumann, Morgenstern, 1944]. For more than fifty years various types of games: two- and multi-persons, discrete- and continuous-strategy, differential, antagonistic and non-antagonistic, co-operative and non-co-operative, one- and multi-level, etc. have been investigated [Dubin, Suzdal, 1981], [Owen, 1968], [Vorob'ev, 1994]. Various areas of game theory applications: in business, technology, military service, etc., have also been investigated. The work [Berzin, 1983] by E.A. Berzin where game theory application to distribution of resources has been investigated is close to our interests. Some aspects of games played in information systems design and maintaining were also mentioned in [Kulikowski, 1990] while in [Kulikowski, 1993] the role of self-organisation in distributed databases development was considered. However, the problem still seems to need deeper investigation. The aim of this paper, being an extended version of the one presented at the International Conference on Computational Science in Cracow (2004), is presentation of a more detailed model of distributed information resources gathering and their profiles forming, based on the theory of *n*-persons games.

2 Basic model assumptions

There will be considered a system of primary *information supply* (*IS*) and a one of *information distribution* (*ID*) as two complementary components of *information market*. The *IS* system is the one where information in the form of various types of documents (electronic, photo, multi-media, hard copy manuscripts, publications, etc.) is offered to the customers. For many years this type of information distribution and exchange was prevailing.

This situation has been changed with computer networks development; the action of *ID* system is based on electronic documents exchange through computer networks. *ID* system thus consists of a set of *open-access data banks* (*OADB*) and of a number of *local data banks* (*LDB*), dedicated to various organisations; the data banks are mutually connected by a computer network, as shown in [<see> Fig. 1].

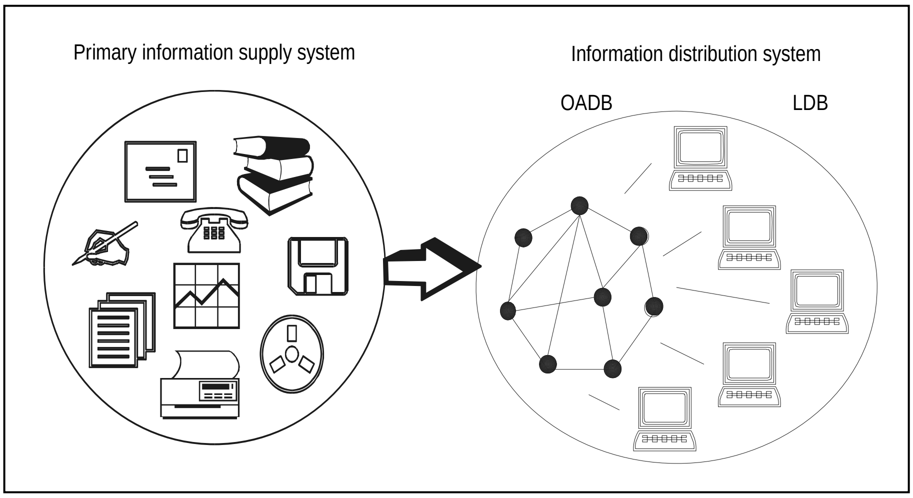

Figure 1: A model of information storage and distribution

Information resources of *OADB*s are permanently extended and supplied with documents from the *IS* system. The role of *OADB*s consists in brokerage: they purchase information documents, select information from them, transform, store and sell or distribute it among the *LDB*s, the last being explored by organisations or by individual users. Information resources of the *LDB*s are thus collected according to the needs of the users. The *LDB*s can be supplied by information not only from *OADB*s, but also from their proper information sources, as well as directly from the *IS* system.

We would not like to go deep into considerations of the mechanisms of *IS* action. For many years a discussion between the adherents of a free world-wide access to information and those of an information market based on commercial principles development has been continued. However, existence of an information market is a real fact and in the below-presented model it will be taken into account. It is not discordant to the fact that certain types of information are accessible free of charge. It is also necessary to remark that information market is based on the categories of information documents' and information services' prices and costs rather than on this of information value. The information value is not easy to define. The price of a computer programming handbook is strongly fixed while value of information contained in it is different for an experienced programmer and for a beginning student of computer science. There is also no direct correspondence between information value and the amount of information. Buying two copies of a computer journal we pay a double price despite the fact that, according to information theory principles, the amount of information in two copies is the same as in a single one. Evidently, the price of books, journals, etc. is not the price of information. Therefore, the notion *information market* is not quite correct: one should be talking rather about a *market of information documents and services.* However, the former notion will be used below for the sake of simplicity.

It is assumed that each document offered at the information market can be described by the following characteristics:

i/ *formal properties*: type of the document, its author(s), title, volume, editor, date of edition, etc.;

ii/ *contents* (described by keywords, descriptors, etc.);

iii/ *acquirement conditions* (name of the seller, price, forms of payment, delivery time, number of available copies of the document, etc.);

iv/ *supply indicator* (number of available copies of the document, if limited, infinite otherwise).

The document characteristics can be formally represented as elements of a *document characteristics space* (*DCS*) being defined as a Cartesian product:

$$DCS = C_f \times C_c \times C_a \times C_s \tag{1}$$

where C_f , C_c, C_a and C_s stand, correspondingly, for the sets of formal, contents, acquirement and supply admissible values of characteristics. We shall denote by $\boldsymbol{h}_i$, $\boldsymbol{h}_i \in DCS$, the characteristic of a document $\boldsymbol{x}_i$, $i \in [1,2,3,\ldots]$.

The state of information market is a time-varying process: at any fixed time-instant t_0, $t_0 \in T$ (T being a real discrete time-axis), the instant-value of the process is given by a finite subset $\boldsymbol{X}(t_0) \subset DCS$ of the documents that are actually available on the market. The members $\boldsymbol{x}_i$ of $\boldsymbol{X}(t)$ thus appear at certain time-instants and next, after a certain time-delay they may disappear. This process of time-varying availability of the documents can also be described by a set of time-functions:

$$\xi_i : T \rightarrow DCS \tag{2}$$

assigning a document characteristic (an element of *DCS*) to each time-instant $t \in T$ so that the components of C_f and C_c remain constant while those of C_a and C_s may be varying in time. The subsets $\boldsymbol{X}(t)$ contain only the elements (documents) for which binary indices δ_i (components of C_a) characterizing the availability of the *I*-th document take the value $\delta_i = 1$. For those documents other components of C_a are also defined, otherwise, if $\delta_i = 0$, to other components of C_a the value $\varnothing$ (= *undefined*) should be assigned. However, the vector $\boldsymbol{\xi}(t)$ consisting of linearly ordered components $\xi_i(t)$ is not known for the past and present time-instants t only. It is reasonable to consider it as an instance of a stochastic vector process $\Xi(t)$ describing the states of the information market changing in time (see [<see> Fig. 2]: available documents are signed by continuous line-segments).

The subsets $\boldsymbol{X}(t)$ determine, at the same time, the areas of possible *instant decisions* of the *OADB* managers. Their aim consists in actualisation and extension of the *OADB*s' resources according to the expected demands of the customers. Therefore, at any t the managers can chose the following decisions: 1^0 to select and to acquire new documents in order to include and to keep them in the data banks, 2^0 to select some documents and to delay a final decision about their acquirement, and 3^0 to reject all actual proposals concerning selling of the documents. In the cases 1^0 an d 2^0 the decision made by the ν-th *OADB* manager ($\nu = 1,2,\ldots,N$, where N denotes the total number of *OADB* managers) takes the form of a subset:

$$\boldsymbol{\psi}_\nu(t) \subseteq DCS \tag{3}$$

whose members correspond to the selected elements of $\boldsymbol{X}$(t) and are such that:

1° the projections of the members of $\psi_\nu(t)$ on $C_f \times C_c \times C_a$ are equal to the corresponding elements of $\boldsymbol{X}(t)$ projected on $C_f \times C_c \times C_a$;

2° the values of the components β_i of C_s in the members of $\psi_\nu(t)$ satisfy the inequalities

$$0 \le \beta_i(t) \le \delta_i(t), \tag{4}$$

which means that the documents acquisition requirement can not exceed the corresponding supply indicators. The document suppliers collect the requirements and try to realise them.

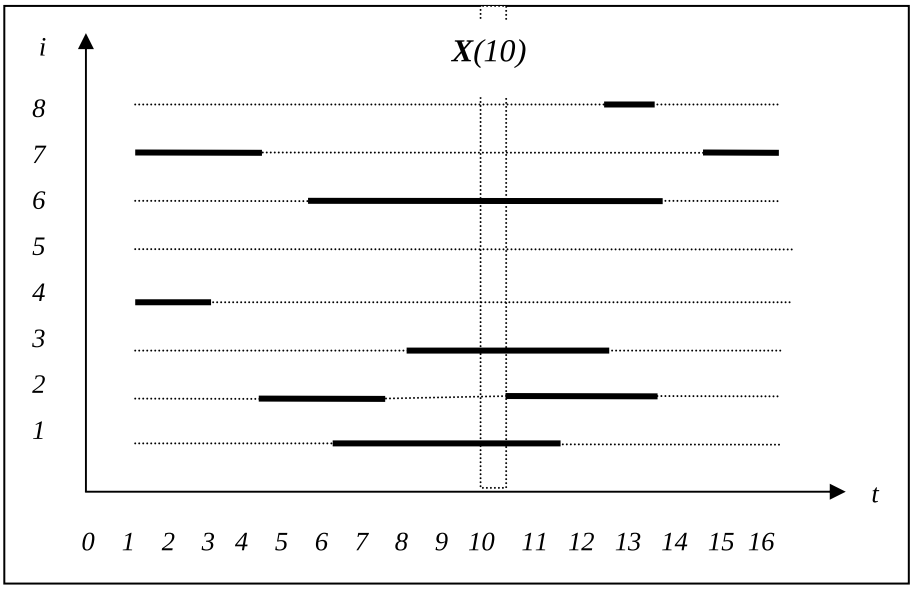

Figure 2: The process of information market states' changes

The number of the copies of the sold document's can not exceed the declared supply indicator. Therefore, it may happen that some document acquisition requirements are not satisfied. However, the strategy of selection of the clients on the information market will not be considered here. In any case, when coming to the next (t+1) time-instant the supply indicators should be actualised: reduced by the numbers of sold document copies and increased by the numbers of the newly supplied ones.

The documents acquired by the *OADB*s can be included into the data banks directly or after a transformation process changing their form or generating some secondary documents on the basis of information selected from the primary ones. In a way similar to the one shown before, a modified *document characteristics space* can be defined:

$$DCS^{(\nu)} = C^{(\nu)}{}_f \times C^{(\nu)}{}_c \times C^{(\nu)}{}_a \tag{5}$$

Here the notions $C^{(\nu)}{}_f$, $C^{(\nu)}{}_c$ and $C^{(\nu)}{}_a$ are, in general, similar to the C_f, C_c and C_a used in (1); however, a dependence on ν indicates that each (ν-th) *OADB* may use its proper language and standards for data files characterisation and define data acquirement conditions for the customers. So, a subset $K^{(\nu)} \subseteq DCS^{(\nu)}$ plays the role of a catalogue of data offered to the managers of *LDB*s or to individual users. A projection of $K^{(\nu)}$ on the document characteristic subspace $C^{(\nu)}{}_f \times C^{(\nu)}{}_c$, $\boldsymbol{L}^{(\nu)} \subseteq C^{(\nu)}{}_f \times C^{(\nu)}{}_c$, will be called a *profile of the* ν-th *OADB* ($OADB^{(\nu)}$), and its elements will be denoted by $\lambda^{(\nu)}$. A substantial difference between the formerly defined characteristics $\boldsymbol{h}_i$ and the characteristic $\lambda^{(\nu)}$ consists in the fact that $\boldsymbol{h}_i$ describes the document in its original form offered on the market while $\lambda^{(\nu)}$ describes any secondary document based on it, available in an electronic form. Forming the profiles $\boldsymbol{L}^{(\nu)}$ of the *OADB*s is the main element of long-term strategies of *OADB*'s managers.

Then, let us take into consideration the *LDB*s' managers' point of view. They represent the interests of some groups of information users (or are information users themselves). The users need to have easy access to information resources suitable for satisfying their intellectual (educational, cultural, etc.) interests or for solving some professional (technological, administrative, etc.) tasks. Let us assign index μ to a certain group of information users. Then their information needs can be formally represented by subsets of a Cartesian product describing an *information requirements space*:

$$IRS^{(\mu)} = C^{*\,(\mu)}{}_f \times C^{*\,(\mu)}{}_c \times C^{*\,(\mu)}{}_a \tag{6}$$

Once more, the notions $C^{*\,(\mu)}{}_f$, $C^{*\,(\mu)}{}_c$ and $C^{*\,(\mu)}{}_a$ are, in general, similar to the C_f, C_c and C_a ones; however, a dependence on μ shows that each (μ-th) *LDB* may use its proper language and standards for data files characterisation and define additional conditions for data acquisition (like admissible cost, time-delay, etc.). In particular, the sets $C^{*\,(\mu)}{}_f$, $C^{*\,(\mu)}{}_c$ and $C^{*\,(\mu)}{}_a$ may contain an element * (*any possible*) to be used if some characteristics of data files or records are not fixed by the users. The subset $\boldsymbol{R}^{(\mu)} \subseteq IRS^{(\mu)}$ characterising information needs of the user(s) can be used in two ways:

1. for information retrieval in $LDB^{(\mu)}$,
2. for actualisation of *local information resources* ($LIR^{(\mu)}$) in $LDB^{(\mu)}$.

$LIR^{(\mu)}$ is also a subset of $IRS^{(\mu)}$. An information retrieval order can be realised if a certain *consistency level* between $\boldsymbol{R}^{(\mu)}$ and $LIR^{(\mu)}$ is reached, otherwise it is necessary to import the necessary data from the *OADB*'s. However, the managers of *LDB*s may conduct a more active policy of users' requirements realisation. This can be reached by a permanent monitoring of the information requirements flow: … $\boldsymbol{R}^{(\mu)}(t\text{-}3)$, $\boldsymbol{R}^{(\mu)}(t\text{-}2)$, $\boldsymbol{R}^{(\mu)}(t\text{-}1)$, $\boldsymbol{R}^{(\mu)}(t)$ (t being the current time) in order to define a preferable *local resources profile*. The last for the given $LDB^{(\mu)}$ can be defined as a subset

$$\boldsymbol{\Lambda}^{(\mu)} \subseteq C^{*\,(\mu)}{}_f \times C^{*\,(\mu)}{}_c \tag{7}$$

such that if $LIR^{(\mu)} = \boldsymbol{\Lambda}^{(\mu)}$ then a considerable part of expected information requirements can be directly realised. The managers of *LDB*s try to achieve this situation within their possibilities as it will be shown below.

3 Strategies for *OADB* managers

The *OADB*s' and *LDB*s' managers are interested in realisation of information requirements of their customers. The managers of *OADB*s are customers on the *IS* market and, at the same time, they are data providers for the managers of *LDB*s. On the other hand, the last ones are data suppliers for the information users. Buying a document $\boldsymbol{x}_i$ available on *IS* market needs covering a cost κ_i being indicated as a component of the corresponding document characteristic. The same document (or data drawn from it) included into the information resources of $OADB^{(\nu)}$ is expected to be distributed among a number of *LDB*s to the information profiles of which it suits. As a consequence, the manager of $OADB^{(\nu)}$ expects to reach proceeds of $r_i^{(\nu)}$. Finally, its expected profit from buying and distributing $\boldsymbol{x}_i$ is

$$c_i^{(\nu)} = r_i^{(\nu)} - \kappa_i \tag{8}$$

It might seem that there is a simple decision rule for the $OADB^{(\nu)}$ manager:

$$\left.\begin{array}{l}\text{purchase } \boldsymbol{x}_i \text{ if } c_i^{(\nu)} \geq e_i^{(\nu)} > 0, \\ \text{postpone the decision if } 0 < c_i^{(\nu)} < e_i^{(\nu)}, \\ \text{do not purchase it otherwise.}\end{array}\right\} \tag{9}$$

where $e_i^{(\nu)}$ is a threshold. However, there arise the following problems: 1st - how to evaluate $c_i^{(\nu)}$ (or $r_i^{(\nu)}$, see (8)), and 2nd - how to fix $e_i^{(\nu)}$?

For answering the first question the following assumptions can be made:

a/ expected proceeds $r_i^{(\nu)}$ can be described by an increasing function of the number of customers whose information profiles $\Lambda^{(\mu)}$ are *consistent* with the information characteristics $\boldsymbol{h}_i$ of $\boldsymbol{x}_i$ and of a total measure of this *consistency*;

b/ the information profiles $\Lambda^{(\mu)}$ of the *LDB*s are not known exactly to the manager of $OADB^{(\nu)}$, he can only approximate them in the form of information profile $\boldsymbol{L}^{(\nu)}$ constructed on the basis of all past and actual information requirements from the *LDB*s;

c/ $r_i^{(\nu)}$ is a decreasing function of other *OADB*s in the *ID* system that will offer $\boldsymbol{x}_i$ or some information drawn from it.

Then, the next problem arises: how to evaluate the *measure of consistency* between a document characteristic $\boldsymbol{h}_i$ and the profile of $\boldsymbol{L}^{(\nu)}$, assuming that (3) holds and $\boldsymbol{h}_i \in C^{(\nu)}_f \times C^{(\nu)}_c$. Let us remark that in general the elements of $C^{(\nu)}_f \times C^{(\nu)}_c$ are not vectors in algebraic sense, but rather some strings of elementary data of various formal nature. Therefore, it is not possible to take an ordinary *distance measure* concept as a basis of a *consistency measure* definition. However, the last can be based on a generalised, *multi-aspect similarity measure* concept proposed in [Kulikowski, 2002]. In this case, if *A* is a non-empty set, then a similarity measure between its elements can be defined as a function:

$$\sigma:\ A \times A \rightarrow [0,1]_c \tag{10}$$

where $[p,q]_c$ denotes a continuous interval between *p* and *q*; here σ is such that:

a/ for each $a \in A$ there is $\sigma(a,a) \equiv 1$;

b/ for any $a, b \in A$ there is $\sigma(a,b) \equiv \sigma(b,a)$;

c/ for any $a, b, c \in A$ there is $\sigma(a,c) \geq \sigma(a,b) \cdot \sigma(b,c)$.

If $\boldsymbol{f}^{(r)} = [f^{(r)}_1, f^{(r)}_{2,\ldots,} f^{(r)}_\gamma, \ldots, f^{(r)}_g]$ and $\boldsymbol{f}^{(s)} = [f^{(s)}_1, f^{(s)}_{2,\ldots,} f^{(s)}_\gamma, \ldots, f^{(s)}_g]$ are two strings of characteristics whose components are of various formal nature then a measure of multi-aspect similarity can be defined as a product:

$$\sigma(\boldsymbol{f}^{(r)}, \boldsymbol{f}^{(s)}) = \sigma_1(f^{(r)}_1, f^{(s)}_1) \cdot \sigma_2(f^{(r)}_2, f^{(s)}_2) \cdot \ldots \cdot \sigma_g(f^{(r)}_g, f^{(s)}_g) \tag{11}$$

This definition can be used directly to the similarity evaluation of documents characteristics. Let $\sigma(\boldsymbol{h}_i, \boldsymbol{h}_j)$ be a similarity measure described on the Cartesian product $\boldsymbol{A} = C^{(\nu)}_f \times C^{(\nu)}_c$. Then it will be said that a member $\boldsymbol{h}_i, \boldsymbol{h}_i \in \boldsymbol{A}$ is *adherent* to a subset $\boldsymbol{L}^{(\nu)} \subseteq \boldsymbol{A}$ on a level ε, $0 < \varepsilon \leq 1$, if there is at least one element $\boldsymbol{h}_j \in \boldsymbol{A}$ such that $\varepsilon \leq \sigma(\boldsymbol{h}_i, \boldsymbol{h}_j) \leq 1$. Here ε is a threshold chosen according to the application requirements.

Adherence of $\boldsymbol{h}_i$ to $\boldsymbol{L}^{(\nu)}$ on a fixed level is a necessary, but not sufficient condition for making a positive decision according to the rule (9). For this purpose the *OADB*$^{(\nu)}$ manager should also take into account: 1[st] - how many characteristics $\boldsymbol{h}_j$ in $\boldsymbol{L}^{(\nu)}$, for all possible $\boldsymbol{h}_j$, satisfy the adherence condition, 2[nd] - how many customers have declared their interests in acquiring the data characterised by $\boldsymbol{h}_j$, and 3[rd] - how long it is passed since the last call for $\boldsymbol{h}_j$. The corresponding, additional data can be stored and included into the information requirement characteristics (see (6)). Taking them into account a *measure of consistency* $\Gamma(\boldsymbol{h}_i, \boldsymbol{L}^{(\nu)})$ between $\boldsymbol{h}_i$ and $\boldsymbol{L}^{(\nu)}$ can be defined as it will be illustrated below.

Let us assume that the contents of documents are characterised by keywords that are presented in a linear order: $w_1, w_2, w_3, \ldots$, etc. Let us take into account two documents, whose characteristics $\boldsymbol{h}_i$, and $\boldsymbol{h}_j$ contain, correspondingly, the subsets of keywords W_i and W_j, There will be considered the sets: $W_i \cup W_j$, and $W_i \cap W_j$, If the cardinal number of a set W is denoted by $|W|$ then the similarity measure of the above-mentioned sets can be defined by a Jaccard similarity measure

$$\sigma(W_i, W_j) = \frac{|W_i \cap W_j|}{|W_i \cup W_j|} \tag{12}$$

This similarity measure can be taken as a basis of the $\Gamma(\boldsymbol{h}_i, \boldsymbol{L}^{(\nu)})$ definition. First, let us remark that $\boldsymbol{L}^{(\nu)}$ can be formally interpreted as a virtual document consisting of all documents whose characteristics were in the past required by a considerable part of customers. Therefore, the formula (12) can be used for evaluation of similarity between $\boldsymbol{h}_i$ and $\boldsymbol{L}^{(\nu)}$, as well. Let us suppose that in the given *OADB*$^{(\nu)}$ the requirements are registered in such a way that to any keyword w_α there are assigned the numbers $e_{\alpha,\tau}$, $\tau = 0,1,2,\ldots,T$ indicating how many times w_α was mentioned in the information calls in the present ($\tau = 0$), as well as in the former time-periods (years). Then the following weight coefficient can be defined:

$$\lambda_\alpha = \sum_{\tau=0}^{T} \frac{e_{\alpha,\tau}}{\tau + 1} \tag{13}$$

and, at last, we can put

$$\Gamma(\boldsymbol{h}_i, \boldsymbol{L}^{(\nu)}) = \sigma(\boldsymbol{h}_i, \boldsymbol{L}^{(\nu)}) \cdot \Sigma_\alpha \lambda_\alpha \quad (14)$$

where the sum on the right side is taken over all keywords occurring both in $\boldsymbol{h}_i$ and $\boldsymbol{L}^{(\nu)}$. The manager of $OADB^{(\nu)}$ thus can assume a proportionality:

$$c_i^{(\nu)} = k \cdot \Gamma(\boldsymbol{h}_i, \boldsymbol{L}^{(\nu)}) \quad (15)$$

(k being a positive coefficient of proportionality) meaning that the greater is the consistency measure between $\boldsymbol{h}_i$ and the profile $\boldsymbol{L}^{(\nu)}$, the higher are the expected proceeds from selling information drawn from $\boldsymbol{x}_i$. The manager of $OADB^{(\nu)}$ thus tries to maximise his gain which is determined by three factors: the information profile $\boldsymbol{L}^{(\nu)}$, the price k'_i of giving access to the document if stored in the resources of $OADB^{(\nu)}$, and the price k''_i of selling the document for a permanent use; the three elements will be called a *current strategy* of the $OADB^{(\nu)}$ manager:

$$\boldsymbol{S}^{(\nu)}_i = [\boldsymbol{L}^{(\nu)}, k'_i, k''_i] \quad (16)$$

According to the fixed $\boldsymbol{S}^{(\nu)}_i$ the manager tries to make his decision about purchasing $\boldsymbol{x}_i$ so that the expected gain $g^{(\nu)}_i$ is maximised.

However, in the situation when the supply indicator of $\boldsymbol{x}_i$ on the market is high (i.e. > 1) the same document can be bought and distributed by other *OADB*s. In such case, assuming that the number of *LDB*s interested in acquiring $\boldsymbol{x}_i$ is unchanged, the incomes of the *OADB*s will be inversely proportional to the number of *OADB*s acquiring $\boldsymbol{x}_i$. The managers of *OADB*s in this case play thus a competitive game with fixed positive total gain. They may all acquire the given document. If they do it at the same time, they will share the gain proportionally to the number of customers requiring access to data from the document at the given *OADB*. Therefore, for each manager it is guaranteed reaching only his maximum gain in the less favourable situation, i.e. when the same document has been acquired by all other managers of the *OADB*s:

$$\overline{g}^{(\nu)}_i = max_{(\nu)}\, min_{(\neq \nu)} \{g^{(\nu)}_i\} \quad (17)$$

the maximum being taken over all possible current decisions in $OADB^{(\nu)}$ and the minimum – over all current decisions made in the other *OADB*s.

However, in the situation when the supply indicator of $\boldsymbol{x}_i$ on the market is high (i.e. >1) the same document can be bought and distributed by other *OADB*s. In such case, assuming that the number of *LDB*s interested in acquiring $\boldsymbol{x}_i$ is unchanged, the incomes of the *OADB*s will be inversely proportional to the number of *OADB*s acquiring $\boldsymbol{x}_i$. Therefore, the managers of *OADB*s in this case play a competitive game with fixed positive total gain. They may all acquire the document; in such case, if they do it at the same time, they will share the gain proportional to the number of customers that will require access to the document. However, the number of *OADB*s purchasing the document a priori is not known. It may thus happen that a certain

OADB acquiring $\boldsymbol{x}_i$ reaches its proceeds much below the expected ones. On the other hand, this number can be evaluated after a time delay (information about newly acquired documents is available to all clients). In such case the manager of $OADB^{(\nu)}$ may, first, postpone his decision in order to recognise the actions undertaken by other players. However, if the decision is made too late then the expected proceeds become low as well.

The formulae (16), (17) describe only current strategies of the $OADB^{(\nu)}$ manager when a given document $\boldsymbol{x}_i$ is considered. For a long period an average strategy of the manager is defined by the information profile $\boldsymbol{L}^{(\nu)}$ and a long-term policy of fixing the prices k'_i, k''_i. If $g^{(\nu)}$ denotes an expected average gain reached by the $OADB^{(\nu)}$ manager under his long-term strategy $\boldsymbol{S}^{(\nu)}$ then his guaranteed long-term gain is given by the expression:

$$\overline{g}^{(\nu)} = max_{(\nu)}\, min_{(\neq \nu)}\, \{g^{(\nu)}\} \tag{18}$$

the maximum being selected from all long-term strategies of the $OADB^{(\nu)}$ manager (including, in particular, the information profile $\boldsymbol{L}^{(\nu)}$), while the minimum is taken over the long-term strategies of all other *OADB*s' managers.

The risk of acquiring documents which will not be needed by the customers may be not acceptable by the *OADB*s' managers. In such case the game rules by establishing a coalition among the managers. The coalition consists in specialisation of information resources of the *OADB*s so that their information profiles $\boldsymbol{L}^{(\nu)}$ are (at least, approximately) disjoint. As a consequence, the *OADB*s will share the information requirements of the customers according to their contents. In such case the expected incomes of the *OADB*s lose the factor of uncertainty connected with the number of competitive *OADB*s offering access to the same documents. The above-described general situation can be illustrated by the following example.

Example 1.

Let us suppose that two *OADB*s are acting on the information market and a new document $\boldsymbol{x}$ appears. The cost if this document is κ and both *OADB* managers use the same function $n(r)$ for anticipating the expected number of customers that may be interested in getting access to the document $\boldsymbol{x}$ when the price of it is r. Usually, $n(r)$ is a decreasing function of r as shown in [<see> Fig. 3].

Both managers can make the same type of decisions: to purchase the document (1) or to reject this proposal (0). The game is symmetrical, as it can be seen in [<see> Tab. 1] indicating the expected gains of both players. However, they do not know each other's decisions.

Therefore, for each player the guaranteed gain is equal ½ $r{\cdot}n(r)$ - κ. It may happen that it is below the threshold level fixed by one or both managers. In such case he (they) should reject the proposal of purchasing the document and his (their) gain will be 0.

M_2 / M_1	0	1
0	0	$r{\cdot}n(r) - \kappa$
1	$r{\cdot}n(r) - \kappa$	$\frac{1}{2}\, r{\cdot}n(r) - \kappa$

Table 1: The guaranteed gains of the players

It can be easily remarked that if the managers act in co-operation, they could make an agreement such that in the above-described situation every second time only one of them purchases the offered document. In such a case his gain is $r{\cdot}n(r) - \kappa$ (while this of the second player is 0). In a long series of decisions they will reach the same average (per a single decision) gain equal ½ $[r{\cdot}n(r) - \kappa]$.

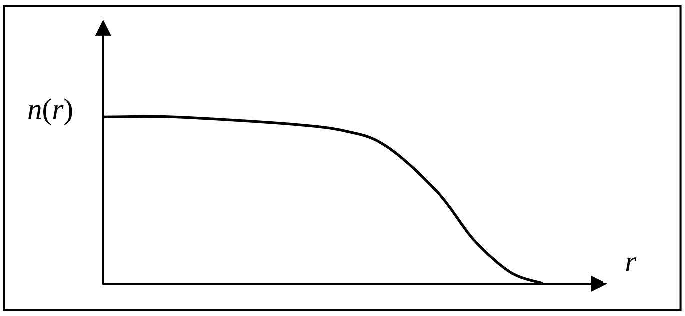

Figure 3: Number of customers interested in getting access to a document as a function of its price

The gain expected by the managers depends on the price r of getting access to the document. However, the product $r{\cdot}n(r)$ determining the gain, as can be seen in [<see> Fig. 4], reaches a maximum at a certain value $r = r^*$. The maximum income of the players thus will be determined by the component ½ $r^*{\cdot}n(r^*) - \kappa$. In any case, they should not fix the price r outside a profitability interval shown below the graph in [<see> Fig. 4].

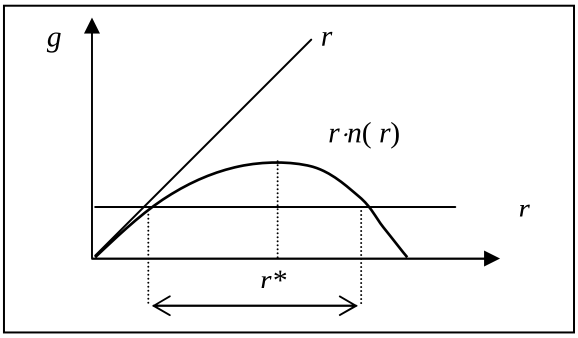

Figure 4: Expected gain as a function of the price r of getting access to a document

The manager can fix neither a lower price (because his incomes could be reduced) nor a higher price (because the number of clients could be reduced).
The managers can also compete for attracting more customers due to offering them better and more effective service, etc. In such case other types of strategies should be considered.
If an agreement between the *OADB* managers is impossible they may try to use randomised strategies. In such case they make their positive (1) decisions

independently and randomly with fixed probabilities. This will be illustrated by the example.

Example 2.

In a situation similar to this described in Example 1 we shall denote by $g_{\alpha\beta}$ the gain reached by the player if his decision is α while the decision of the second player is β, where $\alpha, \beta \in \{0,1\}$. The mean gains are given by [<see> Tab. 2]:

	For the first player	For the second player
g_{00}	0	0
g_{01}	0	$p(1-p)[r \cdot n(r) - \kappa]$
g_{10}	$p(1-p)[r \cdot n(r) - \kappa]$	0
g_{11}	$p^2[½\, r \cdot n(r) - \kappa]$	$p^2[½\, r \cdot n(r) - \kappa]$

Table 2: The average gains of the players

The average gain (per a single decision), the same for both players, is given by:

$$g = p[r \cdot n(r) - ½\, p \cdot r \cdot n(r) - \kappa]$$

Considered as a function of the probability p this function takes the form shown in [<see> Fig. 5]. It is interesting that it takes a maximum at

$$p^* = 1 - \frac{\kappa}{r \cdot n(r)}$$

Let us remark that this probability is not obviously equal 1/2 as it could be assumed at the first glance.

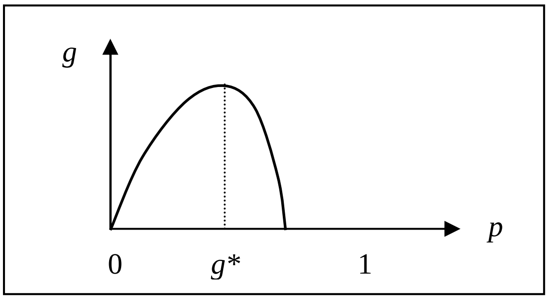

Figure 5: Average gain as a function of probability p of making a decision of purchasing a document

The maximum average gain calculated for $p = p^*$ is given by the following formula:

$$g_{\max} = \frac{1}{2} r \cdot n(r) + \frac{\kappa^2}{2r \cdot n(r)} - \kappa$$

4 Remarks on the strategies for *LDB* managers

The relationships between the *LDB* managers and the information users is not the same as this one between the *OADB*s' and *LDB*s' managers. The difference consists in the fact that the *LDB*s' managers and their information users usually act within the same organisations Therefore, up to a certain degree, their interests are convergent. However, a source of conflicts may be connected with the fact that a *LDB*s' manager has at his disposal a limited amount of financial resources that can be used for the realisation of information requirements of his customers. In such case he should establish a system of priorities in acquiring new data from the *OADB*s. As a consequence, the users attached to a certain $LDB^{(\mu)}$ may have partially common and partially competitive interests. Assuming that the *LDB*s' manager has no reason for distinguishing some information users against some other ones he will try to make his decisions on the basis of the information profile $\Lambda^{(\mu)}$. His problem is thus as follows:

1. he has at his disposal an amount ζ of financial funds that can be spent for extending the $LIR^{(\mu)}$ for a certain time T^*;
2. he knows the actual information requirements $\boldsymbol{R}^{(\mu)}$ of his customers (data users) that can not be realised on the basis of $LIR^{(\mu)}$;
3. he knows the actual information resources of the *OADB*s as well as the corresponding data selling and/or data access conditions.

How to realise the information needs of the clients within the financial and/or formal limits? It is assumed that the prices of an incidental access to some data and of data purchasing for their permanent use within the organisation are different (the last being usually much higher). Therefore, if an analysis of the *OADB*s' resources shows that a certain requirement can be realised by importing data then it is necessary to determine the expected costs of: *a/* data purchasing and *b/* data access during the time-interval T^* for various *OADB*s and to find out the minimum ones. At the next step the conflict of interests of various customers should be solved: there has been determined a set of data acquisition offers: $O_1, O_2, O_3,\ldots$ etc., each one being characterised by its expected cost χ_ι and by the subset of clients that might be interested in using the given data in the time-interval T^*. The problem is: how to select a subset of the offers for their final acceptation?

This problem can be solved if a system of ordering the offers is defined. Each offer can be characterised by a vector whose components indicate: 1° the expected cost χ_ι, 2° the number b_ι of clients that are interested in using the given data, and 3° the mean measure of similarity s_ι of the offered data to the actual (and expected, if possible) data requirements. Therefore, the problem arises of semi-ordering of the vectors $\boldsymbol{u}_\iota = [\chi_\iota, b_\iota, s_\iota]$ in a three-dimensional space. The problem can be solved easily on the basis of *K*-space (Kantorovich space) concept [Kantorovich, Vulich, Pinsker, 1959]. Then the strategy of the *LDB* manager consists in rearranging the offers: $O_{\rho 1}$, $O_{\rho 2}$, $O_{\rho 3}$, etc. (ρ_1, ρ_1, ρ_1, etc. being some integers) and in accepting realisation of as many offers as possible within the given financial funds.

Finally, let us take into account the relationships among the group of *OADB*s and of *LDB*s managers. The conflict among them consists in the fact that the first ones try to establish the data delivery prices as high as possible, while the second ones can spent only a limited amount of financial funds trying to achieve as much information as possible. However, if the prices of data are too high, the customers may reduce their information requirements, and in such case the total amount of financial funds for data purchasing will be decreased. Therefore, there is a point of maximum data prices leading to maximum incomes of the data suppliers. The managers of *LDB*s may influence the data suppliers in order to reduce the costs when they act in coalition.

5 Conclusions

We tried to show that the relationships between the managers of *OADB*s and of *LDB*s have the form of a *n*-person game of partially co-operative, partially non-co-operative type with incomplete information. In the most typical situations it is possible to define the strategies of the players. The strategies are realised by creation of database profiles that are used in making decisions concerning documents or data acquisition. The decision rules of the players can be strongly optimised only in particular, rather simple cases. In general, it can be shown that a co-operation between the database managers may lead to less risk in reaching their goals. However, many other problems still have to be investigated.

Acknowledgements

I thank the reviewer of this paper for his valuable remarks. I also thank Ms Renata Maksymowicz for her help in making stylistic amendments in the paper.

References

[Berzin, 1983] Berzin E.A.: Optimalnoe raspredelenije resursov i teorija igr (in Russian: Optimum Distribution of Resources and Theory of Games). Radio i svjaz', Moskva (1983).

[Dubin, Suzdal 1981] Dubin G.N., Suzdal V.G.: Vvedenije v prikladnuju teoriju igr (in Russian: Introduction to Applied Theory of Games). Nauka, Moskva (1981).

[Kantorovich, Vulich, Pinsker, 1959] Kantorovich L.V., Vulich B.Z., Pinsker A.G.: Funkcjonalnyj analiz v poluuporiadocennych prostranstvach (in Russian: Functional analysis in semi-ordering spaces). GITTL, Moskva (1959).

[Kulikowski, 1990] Kulikowski J.L.: Problemy budowy systemów informatycznych w środowiskach społecznych (in Polish: Problems of Information Systems Design in Social Environment). Techniki i metody rozproszonego przetwarzania danych, cz. V (M. Bazewicz ed.). OW Politechniki Wrocławskiej, Wroclaw (1990).

[Kulikowski, 1993] Kulikowski J.L. Self-organization in Distributed Data Bases. Information Systems' Architecture and Technology ISAT'93 (M. Bazewicz ed.). Wydawnictwo Politechniki Wrocławskiej, Wroclaw (1993), 94-104.

[Kulikowski, 2002] Kulikowski J.L.: From Pattern Recognition to Image Interpretation. Biocybernetics and Biomedical Engineering, vol. 22, No 2-3 (2002), pp. 177-197.

[Neumann, Morgenstern, 1944] Neumann von J., Morgenstern O.: Theory of Games and Economic Behavior. Princeton (1944).

[Owen, 1968] Owen G.: Game Theory. W.B. Saunders Company, Philadelphia (1968).

[Vorobyev, 1994] Vorob'ev N.N.: Osnovy teorii igr. Beskoalicjonnye igry (in Russian: Principles of games theory. Non-co-operative games).Nauka, Moskva (1994).