Modeling of information-analytical support for the timber industry complex

Alexander Poluektov[1]*, *Tatiana* Skvortsova[1], *Evgeny* Anikeev[1], and *Alexander* Yagodkin[1]

[1] Voronezh State University of Forestry and Technologies named after G. F. Morozov, Voronezh, Russia

Abstract. The article describes the methods of designing the information network of the timber industry complex. Algorithms for generating user requests and converting them to a format that is understandable for processing in the system are described. A model of the behaviour of a multi-purpose information system is developed and mathematically justified, and its reachability is checked. The resource allocation problem is also formulated and the solution of the problem is mathematically justified. The possibility of conflicts appearing in the system is studied and evaluated, possible conflicts are identified and a mathematical apparatus for resolving conflicts is developed, as well as subsets of desirable and undesirable behaviour of subsystems are studied. In addition to the theoretical substantiation of the design methodology, the article also contains mathematical tools that can be useful for developing various information networks.

1 Introduction

The design of the information network of the timber industry complex is based on the following algorithm: first, a request is generated from the user, which is converted into a format understandable for processing in the system, then the input object is converted into an output object and, finally, the output documentation is generated. Based on the specified algorithm, the modelling of the general information technology model is performed, which is built as a set of the following models:

query sets S_x are represented as a null graph G_x xcomposed of query vertices $N(X) = (N_1(X_1), N_2(X_2), \dots, N_m(X_m))$;

interactions of multiple requests S_{NX}.

system for converting an input object to an output object S_{XY}. It is represented as a mathematical functionи F defined on the set of the input data of the object X and the product by an element from the sets a C of the system states, the result will be a set of sets of the output object Y;

a relationship system that defines for each element from the output Y co a corresponding document from the sets of output documents S_{YN}.

*Corresponding author: palv2006@yandex.ru

The output document sets S_Y are defined in the same way as the set S_x in the form of a null graph with the set of vertices $N(Y) = (N_1(Y_1), N_2(Y_2), \dots, N_m(Y_m))$.

To understand the algorithms for constructing models, it is necessary to conduct a detailed analysis of each of them [4].

2 Materials and methods

To build models and S_{XY} , you need to set a system for converting the input object into the output object. To do this, we will set the graph $\breve{G}$ in the form of trees, which describes the interaction of the subsystem Γ_0 with other subsystems of the information network.

We determine the reachability of the subsystem Γ_j with graph G_J from an information resource from the subsystem Γ_0 with graph G_0 if there are non-empty subsets of input and output objects such that the subset of output objects Y is the reachability set of the subset of input objects X. Let's introduce the reachability relation $V_0 \to V_j$.

Let's consider the matrix of communication on the reachability of information resources Γ_0 with the subsystem Γ_j. And also let's build the connection matrix $Z_{X_0 Y_j}$, where $j = \overline{1, n}$. To do this, for each $X_{0k} \subset X_0$, we construct the reachability sets Y_j and find their union.

3 Results and discussion

Let's construct a model of the behaviour of Γ_0 as a multi-purpose information system. To do this, we define a set of goals $N(X) = \{N_1(X_1), N_2(X_2), \dots, N_m(X_m)\}$, where X_i is the set of states of the specified format. To do this, describe constraints on the vectors X_i and specify the format N_i, as well as construct vectors for evaluating $P(N_i)$ properties of the goal.

Let's take a closer look at how you can set the conditions for achieving the goals N_i to do this, you need to consider what different strategies E_i exist in the system, for example: neutral, cooperative, and competing. If the number of strategies is greater, we will conditionally assume that their number is about m and to select the desired strategy, we will set the distribution function $\varphi(e)$ which takes the values $1, 2, \dots, m$ and put N_i its Dirichlet domain in correspondence $\mathrm{Э}_i$:

$$\mathrm{Э}_i(E) = \{e_k \in E_i : \varphi(e_k) = i\}, i = \overline{1, m} \tag{1}$$

The condition that all goals are achievable is

$$\bigcup_{i=1}^{m} \mathrm{Э}_i(E) = E \tag{2}$$

If the Dirichlet domain for $i \neq j$ does not intersect, then the problem of choosing strategies does not exist. If $\varphi(e)$ is an ambiguous function, then by introducing the Cartesian product $\mathrm{Э}(E) = \mathrm{Э}_1(E) \times \mathrm{Э}_2(E) \times \dots \times \mathrm{Э}_m(E)$, we reduce the problem to the condition of defining the set E, for which the choice of strategies $e_{ij} \in \mathrm{Э}_i(E)$ is made one at a time.

The strategy selection algorithm can be represented as [2]:

- set $E(e)$ such that $\forall\, i = \overline{1, m}$ и $\forall\, j = \overline{1, |\mathrm{Э}_\iota|}$
- there is a strategy $e_{ij} \in \mathrm{Э}_i(E)$ such that (3)
- $\Phi(E(e)) = min$,

where $\Phi(E)$ is the utility function.

Let us consider the formulation of the problem of resource allocation R. Let R be a matrix of resource assignments. We introduce a resource assignment matrix $R_k(E, X) = \{t_i^k \delta_i\}$, where t_i^k is the amount of resource r_k required to implement the goal N_i when choosing k-th strategy e_{ik}, $\delta_i = 1$ if the resource is assigned and $\delta_i = 0$ otherwise. Then the tasks are set as follows:

$$\text{find } R_k(E, X) \text{ such that } \Phi(R_k) = min \tag{4}$$

Based on the considered arguments, we obtain a general optimization model that can be defined in the form of a vector optimization problem [1]:

$$P(X) \underset{X \subset D}{\rightarrow} Opt, D: P(X) = P(P(E, X), P(K)) \tag{5}$$

where Opt is an operator that implements one of the principles of vector optimization.

Let each of the subsystems Γ_j in the process of functioning strive for its local goal N_i, and for each of the subsystems we can define utility functions ψ_j. We will set for the subsystem of the center of Γ_0 expected utility $\psi_0^{\text{ож}} = \psi_0(z_0)$ where $z_0 \in \{x_0^j\}$ and the real usefulness of $\psi_0^p = \psi_0(z_0)$, where $z_0 \in \{Y_0^j\}$, determined as a result of his cooperation for the implementation of z_0 with j-th subsystem.

Let's define the derivative with respect to the system:

$$\psi_0^{p\prime}(j) = \frac{\partial \psi_0^p}{\partial \Gamma_j} = \lim_{\Delta z_0 \to 0} \frac{\psi_0(z_0 + \Delta z_0) - \psi_0(z_0)}{\Delta z_0} \tag{6}$$

where $\Delta z_0 = z_0 - z_j$, $\psi_0(z_0 + \Delta z_0) = \psi_0^p$.

In the process of achieving their local goals, Γ_0 and Γ_j, three possible relationships arise:

conflict relation: the subsystem Γ_j conflicts with the center Γ_0 if $\psi_0^p < \psi_0^{\text{ож}}$ or $\psi_0^{\text{ож}\prime}(j) < 0$.

This occurs when the functioning of the subsystem Γ_j to search for the necessary information leads to a decrease in the usefulness of the center Γ_0.

co-operation relation: the subsystem Γ_j contributes to the center Γ_0 if $\psi_0^p > \psi_0^{\text{ож}}$ or $\psi_0^{\text{ож}\prime}(j) > 0$.

This occurs when the functioning of the subsystem Γ_j to search for the necessary information leads to an increase in the usefulness of the center Γ_0.

3. indifference relation: a subsystem Γ_j is indifferent to the center Γ_0 if $\psi_0^p = \psi_0^{\text{ож}}$ or $\psi_0^{\text{ож}\prime}(j) = 0$.

This occurs when the subsystem Γ_j agrees with the expected utility of the center Γ_0.

To eliminate conflicts, we distinguish binary relations of conflict, assistance and indifference between all pairs Γ_0 and Γ_j, j= $\overline{1, n}$, as well as subsets of desirable and undesirable behaviour of subsystems.

Then we construct the utility function $\Phi(Z)$ of the entire system, as well as a set of local real utility functions $\{\psi_j^p(z_j)\}$, $j = \overline{1, n}$, which form so-called partial conflict with the utility function of the system [3]. This will allow you to create and solve an optimization problem of the form

$$P(Z) = (\Phi(Z), \psi_1^p(z_1), \dots, \psi_n^p(z_n) \underset{z \in Z}{\rightarrow} Opt \tag{7}$$

It can be reduced to a sequence of lower-dimensional problems:

$$\left(\Phi(Z), \psi_j^p(z_j)\right) \underset{z_j \in Z_j}{\rightarrow} Opt, j = \overline{1, n} \tag{8}$$

Since Γ_0 must approach the maximum benefit for itself and we can assume that $\Phi(Z) = \psi_0^p(z)$. In this case, the utility function of the system Γ_0 establishes a strict linear order $\hat{L}_j = \hat{L}_j$ (>Φ) on the set Z_j.

Let us define qualitative orders for the center Γ_0 по полезности $\Phi(Z_j)\ \tilde{L}_j = \tilde{L}_j\ (\succ \Phi)$ and subsystem Γ_j in terms of utility $\psi_j^0(z_j)\ \tilde{l}_j = \tilde{l}_j(\succ \psi_0^p) = l_j(\psi_j^p)$.

When considering subsystem Γ_j to the behaviour of Γ_0 two options may arise:

1. $\check{Z}_j = \{\max L_j\}$ and $L_j = l_j$, so $\Phi'(z_j) > 0$ and $\psi_j^{p'}(z_j) \geq 0\ \forall z_j$, i.e. the center Γ_0 chooses the solution $z_j^* = \max\{\tilde{L}_j\}$ и $\psi_j^p = \Phi(z_j^*)$.

2. $\check{Z}_j \neq 0$ и $\tilde{L}_j \cap \tilde{l}_j \neq \tilde{L}_j$, so $\Phi'(z_j) < 0\ \forall z_j$, i.e. the solution from the set $\check{Z}_j$ is the best compared to solutions that do not belong to this set.

Let us estimate the behaviour of the center Γ_0 under the conditions of interaction with the subsystems Γ_j, j= $\overline{1,n}$.

The guaranteed win function for Γ_0 is:

$$\chi = \sum_{\Gamma_j \overline{>I} \Gamma_0} \alpha_j(\psi_0^p - \psi_0^0) + \sum_{\Gamma_j > I \Gamma_0} \beta_j(\psi_j^p - \psi_0^0) \tag{9}$$

where α_j, β_j are the weights of subsystems in the system that define the properties of relationships between systems.

$$0 \leq \alpha_j, \beta_j \leq 1, \sum_{\Gamma_j \overline{>I} \Gamma_0} \alpha_j = 1, \sum_{\Gamma_j > I \Gamma_0} \beta_j = 1 \tag{10}$$

Solutions, under the conditions of interaction Γ_0 with Γ_j, can be chosen from the sets a $Z_{\text{пер}} \subset Z$, for which the inequality is fulfilled:

$$\sum_{j=1}^{n}(\psi_0^p - \psi_0^0) \geq \chi \tag{11}$$

Having solved the covering problem, we can find a solution from the set $Z_{\text{пер}}$:

$$\rho(X_0, Y_0) \underset{y_j \in Y_j}{\rightarrow} min$$

$$Y_j = \delta_0^j Y_j^0, Y_0 = \bigcup_{j=1}^{n} \delta_0^j y_i \tag{12}$$

$$\sum_{j=1}^{n}(\psi_0^p - \psi_0^0) \geq \chi \tag{13}$$

where $p(X_0, Y_0)$ is the distance between X_0 and Y_0,

$$\delta_0^j = \begin{cases} 1, \text{if } \Gamma_0 \text{ forms an information array through } \Gamma_j \\ 0, otherwise \end{cases} \tag{14}$$

The resulting problem does not have an efficient solution algorithm and can be solved on the basis of random search with local optimization.

We study a method that allow set us to estimate the components of the weight vectors α and β [5, 6].

Consider an isomorphic transformation $\Gamma = \left(\{\Gamma_j\}_N, G = G(\Gamma, E)\right) \Leftrightarrow \Omega = (\{\Omega_j\}_N, G = G(\Omega, \mathrm{E}))$ in which the graph $G(\Gamma, E) = G(\Omega, E)$ is given by the set of vertices $\{\Gamma_j\} = \{\Omega_j\}$ and the set of arcs $E = \{e_i\}$ defines partial order $> I_{\text{н}}$ We will strive to ensure, that the information system in the process of functioning will strive to achieve its goal, and the utility is equal ψ_j^p.

We define from the set Ω subsets of elements $\Omega^{>I}$ and $\Omega^{\overline{>I}}$ for which the binary relations $> I$ and $\overline{> I}$. Let us define binary relations $> I, \overline{> I}$, $> I_{\text{н}}$ with adjacency matrices $\underset{n\times n}{\mathrm{A}} = \{a_{ij}\}$, $\underset{n\times n}{\mathrm{A}^c} = \{a_{ij}^c\}$, $\underset{n\times n}{\mathrm{A}^{\text{н}}} = \{a_{ij}^{\text{н}}\}, n, m \leq N$, whose elements are 1 if the relation holds, and 0 in otherwise.

Let us consider the following values: $\varepsilon_{\text{вых}}(\Omega_i)$ is the half-degree of the vertex exit Ω_i which determines the number of arcs leaving the vertex Ω_i, and $\varepsilon_{\text{вх}}(\Omega_i)$ is the half-degree of the vertex entry Ω_i, which determines the number of arcs entering the vertex Ω_i:

$$\varepsilon_{\text{вых}}(\Omega_i) = \sum_{i=1}^{r} \omega_{ij}, \ \varepsilon_{\text{вх}}(\Omega_i) = \sum_{j=1}^{r} \omega_{ij} \tag{15}$$

where $\{\omega_{ij}\} = (A \vee A_c \vee A_{\text{н}})$

If $\varepsilon_{\text{вых}}(\Omega_i) = 0$, then the action of Ω_i does not affect the formation of the relation $\gtrsim I$, and the $\varepsilon_{\text{вых}}(\Omega_i) > 0$ determines the number of subsystems that Ω_i affects when forming the relation $\gtrsim I$.

If $\varepsilon_{\text{вх}}(\Omega_i) = 0$, then the element Ω_i affects the relation $\gtrsim I$, and $\varepsilon_{\text{вх}}(\Omega_i) > 0$ determines the number of subsystems that affect Ω_i when forming the relation $\gtrsim I$..

If $\varepsilon_{\text{вых}}(\Omega_i) = \varepsilon_{\text{вх}}(\Omega_i) = 0$, then the vertex is isolated, which indicates errors in the formation of the structure of the information network Ω.

We get

$$\varepsilon(\Omega_i) = \frac{\varepsilon_{\text{вых}}(\Omega_i)+\varepsilon_{\text{вх}}(\Omega_i)}{\sum_{i,j+1}^{r} \omega_{ij}} \tag{16}$$

evaluates the weight of the subsystem Ω_i under conditions of relations $\gtrsim I$ between subsystems of the set Ω.

Consider the properties of the conflict $> I$ and consent $\gtrsim I$ relations with matrices A and A_c, accordingly. Let us assign to each arc e_{ij} of the graph G weight s_i from the set of weights $S = \{s_{ij} = |\psi_j'(\Omega_i)|\}$. Then Ω_i is put in the correspondence

$$\varepsilon(\Omega_i) = \frac{\varepsilon_{\text{вых}}(\Omega_i,S)+\varepsilon_{\text{вх}}(\Omega_i,S)}{\sum_{i,j=1}^{n} s_{ij}} \tag{17}$$

where $\varepsilon_{\text{вых}}(\Omega_i, S) = \sum_{i=1}^{n} s_{ij}\, a_{ij}$, $\varepsilon_{\text{вх}}(\Omega_i, S) = \sum_{j=1}^{n} s_{ij}\, a_{ij}$.

Formula (16) sets an estimate of the rank of subsystems in the system, taking into account the weights S, and formula (17) – without taking into account the weights.

We have the following weight estimates:

$\alpha_j = \varepsilon(\Omega_j)$, $j = \overline{1, N}$ for the entire information system as a whole;

$$\alpha_j = \varepsilon(\Omega_j), j = \overline{1,m} \text{ over the kernel of agreement } \Omega^{\overline{>I}} \tag{18}$$

$\beta_j = \varepsilon(\Omega_j)$, $j = \overline{1, n}$ along the core of the conflict $\Omega^{>I.}$.

Or

$$\alpha_j = \frac{\varepsilon(\Omega_j)}{\varepsilon(\Omega)} \text{ over the entire system } \Omega \text{ together with } \Omega^{\overline{>I}} \tag{19}$$

$\beta_j = \frac{\varepsilon^{\overline{>I}}(\Omega_j)}{\varepsilon(\Omega)}$ along the agreement kernel $\Omega^{\overline{>I}}$ together with S,

where $\varepsilon(\Omega) = \sum_{\Omega_j \in \Omega} \varepsilon(\Omega_j) + \sum_{\Omega_j \in \Omega^{\overline{>I}}} \varepsilon^{\overline{>I}}(\Omega_j)$.

Or

$$\alpha_j = \frac{\varepsilon(\Omega_j)}{\varepsilon(\Omega)} \text{ over the entire system } \Omega \text{ together with } \Omega^{>I} \tag{20}$$

$\beta_j = \frac{\varepsilon^{>I}(\Omega_j)}{\varepsilon(\Omega)}$ along the kernel and $\Omega^{\overline{>I}}$ together with S,

where $\varepsilon(\Omega) = \sum_{\Omega_j \in \Omega} \varepsilon(\Omega_j) + \sum_{\Omega_j \in \Omega^{\overline{>I}}} \varepsilon^{\overline{>I}}(\Omega_j)$.

Or

$$\alpha_j = \frac{\varepsilon^{>I}(\Omega_j)}{\varepsilon(\Omega)} \text{ by the agreement kernel } \Omega^{>I} \text{ together with } \Omega^{\overline{>I}} \quad (21)$$

$\beta_j = \frac{\varepsilon^{\overline{>I}}(\Omega_j)}{\varepsilon(\Omega)}$ along the conflict kernel $\Omega^{\overline{>I}}$ together with $\Omega^{>I}$,

where $\varepsilon(\Omega) = \sum_{\Omega_j \in \Omega^{>I}} \varepsilon^{>I}(\Omega_j) + \sum_{\Omega_j \in \Omega^{\overline{>I}}} \varepsilon^{\overline{>I}}(\Omega_j)$.

The decision maker performs the final construction of estimates according to formulas (18) - (21).

The algorithm for the implementation of information technology for the search and formation of necessary information consists of the stages of formation:

1. requests for information search.
2. a multi-purpose user request.
3. the input object of the X system.
4. Systems for converting input object X to output object Y:
- reachability graph;
- information space of system interaction;
- winning functions based on the conflict interaction of subsystems;
- search for the necessary information vectors.
5. lots of output documents.
6. the user's document.

Previously, we была built a model of information technology search, consisting of a set of five models [3]. Let's consider the model of communication between a set of queries and the input S_{NX}. It establishes a relation between the set of queries $N(X)$ of the model S_x and the set of the input object X of the system Γ_0. And the condition must be met

$$X = \cup_i X_i \quad (22)$$

The query model is defined as a set of the set of initial data structures $N_{\text{исх}}(X)$, the set of derived data structures $N_{\text{пр}}(X)$, and the set of organizational and technological constraints and data properties $M(N)$: $N_i(X_i) = \{N_{\text{исх}}(X_i), N_{\text{пр}}(X_i), M(N)\}$. The result of query is represented by an array of information X_i of the specified format N_i.

The problem arises of constructing conditional unions (taxa) of data, under which condition (22) is met. Let us denote the set of taxa $T = \{T_1, T_2, \dots, T_l\}$.

To assess the elimination of incompleteness неполноты or unreliability of information, it is necessary to simulate the searcha for information based on user requests, which will lead to the formation of tasks and the selection of the optimal taxon $T_{\text{опт}}$ that ensures the fulfilment of condition (22).

Let's explore possible solutions to these problems. When studying a user query a $N(X)$, you need to evaluate what properties it has $M\ (N)$. In order to evaluate the properties of each query for Γ_0 you need to find the efficiency vector:

$$Ef(M, X) = \{Ef_1, Ef_2, \dots, Ef_q\} \quad (23)$$

Since the initial information about the query parameters $N\ (X)$ is incompleteтy and uncertain, we pay additional attention to its formalization. There are several main types of information granules used for generating initial information.

The description of N_i available to Γ_0 can be represented as a vector of information granules, with which the Γ_0 expert can form a description of each element of the set of queries:

$$M_j = \{\mu_1, \mu_2, \dots, \mu_\nu\} \quad (24)$$

The Γ_0 system selects one of the merge options based on the available information about the properties and parameters of the query $N_i(X_i)$. That is, if, for example, all requests have approximately the same deadlines, then the time criteria are not considered, and the combination is performed according to other properties (cost, functional purpose, etc.).

To define the set of conditional combinations of user requests, you can use, for example, the method of selecting alternatives based on a generalized criterion (if the criteria equivalent) or a lexicographic method (if the criteria are ordered by importance).

We will use the first method, improving it somewhat by using interval estimates of λ for the formation of taxa.

Let each of the criteria Ef_q correspond to a set of values D_{iq} that satisfy the selection conditions. Thus, each query $N_i(X_i)$ corresponds to the value D_{iq}, for each of the criteria Ef_q. Then conditional joins are searched for using the following algorithm:

A score is defined for each criterion Ef_q

$$D_{q_{max}} = \max_q D_{qi} \tag{25}$$

We determine the normalized scores of each query $N_i(X_i)$ based on the criteria

$$\breve{D}_{qi} = \frac{D_{qi}}{D_{q_{max}}} \tag{26}$$

For each user request, we calculate the minimum criterion score $D_{i\,min} = \min_q \breve{D}_{qi}$.

We define the degrees of similarity of queries λ in the form of the interval $\lambda = [\lambda_1, \lambda_2] \subset [0,1]$, where $\lambda_1 < \lambda_2$. The taxon T_y is formed from N_i, for which the condition $\lambda_1 \leq D_{i\,min} \leq \lambda_2$ is satisfied for the given degree of similarity.

Changing the values of the left λ_1 and right λ_2 boundaries of the degree of similarity, we get a set of conditional unions.

Since the weight of criteria can vary in practical tasks, it is advisable to build a set of conditional associations according to the hierarchical principle. We will use the lexicographic search method. First, we select the set of information queries with the highest score for the most important criterion, and then select those queries that have the best score for the next criterion from the ordered ones.

Let's order the criteria Ef_q by importance: $Ef_1, Ef_2,\ldots$..

Then the decision maker selects the level $\gamma \in [0,1]$, for which the set of best queries is determined.

To do this, we define the lower and upper bounds of subsets for evaluating the query N_i by the criteria under consideration:

$$\underline{D}_{qi} = \inf_{D_{qi} \geq \gamma} t, \overline{D}_{qi} = \sup_{D_{qi} \geq \gamma} t \tag{27}$$

where t is the value of the i-th request D_{qi} of the γ -level.

For two requests of information N_i and N_{i+1}, given in a random way, indicators of mutual excess of criterion ratings are calculated:

$$\lambda\left(D_{qi} > D_{qi+1}\right) = \frac{\overline{D}_{qi} - \overline{D}_{qi+1}}{\overline{D}_{qi} - \underline{D}_{qi}} \tag{28}$$

$$\lambda\left(D_{qi} < D_{qi+1}\right) = \frac{\underline{D}_{qi+1} - \underline{D}_{qi}}{\overline{D}_{qi} - \underline{D}_{qi}} \tag{29}$$

The indicator of whether *the* i-th query belongs to the best set is calculated:

$$\nu_{qi} = sup\left\{0, \max_{D_{qi}, D_{qi+1}} \lambda\left(D_{qi} > D_{qi+1}\right) - \max_{D_{qi}, D_{qi+1}} \lambda\left(D_{qi} < D_{qi+1}\right)\right\} \tag{30}$$

If $\nu_{qi} \geq \gamma$ for several alternatives, then the level γ is selected again and the process is repeated from the beginning.

By changing of the value γ, we can obtain a set of taxa T.

Previously, we constructed a model of the optimal behaviour Γ_0 under conditions of interaction with the external environment [2]. The proposed algorithm allows us to find the values of the utility function Φ_i for each request $N_i(X_i)$. The task arises to determine the taxon whose elements will have the maximum level of utility.

Then

$$T_y = \frac{\cup_i N_i(X_i).}{\Phi_i(Z)} \tag{31}$$

The utility function

$$\Phi_{T_y} = \cup_i \Phi_i(Z), \forall N_i \in T_y \tag{32}$$

$$\Phi_{T_y} = \frac{\sum_i \Phi_i(Z)}{|T_y|} \tag{33}$$

where $|T_y|$ is the power of the taxon.

if the union, provides the maximum utility function, then it will be optimal

$$max\Phi_{T_y} \tag{34}$$

the optimal taxon is calculated using the formula

$$T_{\text{опт}} = \frac{N}{max\Phi_{T_y}} \tag{35}$$

4 Conclusion

An algorithm for constructing an optimal query pooling:

1. defining criteria (23).
2. utility functions are defined for each query.
3. a set of conditional taxa is constructed according to condition (22).
4. the value of the utility function is calculated for all T_y of the set N (31).
5. the optimal taxon (34) is determined from the condition of the maximum of the utility function (33).

References

1. A.V. Podmazov, Student Bulletin **2-5(147),** 58-61 (2021).
2. A.N. Skoba, V.K. Mikhailov, S.A. Nazarov, A.N.A. Ayesh, Engineering Bulletin of the Don **2(53),** 15 (2019).
3. D. Kan, D. O. Tsaplin, R. O. Ivlev [et al.], Technologies of the future: proceedings of the VII international scientific and technical conference of students and postgraduates, Moscow, June 22-26, pp. 60-68 (2023).
4. V. A. Mokritsky, *Problems of information security*: V All-Russian scientific and practical conference with international participation, Simferopol - Gurzuf, February 14-16, 2019. V.I. Vernadsky Crimean Federal University, Simferopol-Gurzuf: IP Zueva T.V., pp. 109-110 (2019).
5. D. E. Solovey, A.V. Skorikov, Information resources of Russia **1(83),** 13 (2005).

6. S.S. Fomin, Innovations based on information and communication technologies **1,** 474-476 (2013).