CERTAIN INTEGRAL PROPERTIES OF INFORMATION SYSTEMS OF HIERARCHIC TYPE

L. S. Kozachkov

UDC 519.2:007

One of the central problems of cybernetics is the study of the general laws of information transformation in complex control systems [1]. Problems of this kind often have to be solved under conditions of incompleteness and indeterminacy of the information about the structure and state of the system. Difficulties arising here, caused by noncorrespondence of the models formulated to the actual system, are overcome to some extent by going over to simpler models which statistically describe the system as a whole, and developing special methods of statistical estimation of its functioning [2].

Below we describe one such model which was formed using an analysis of systems of scientific information flows, certain properties of a language and a series of statistical distributions [3, 4, 6].

However, certain results here apparently can be used when analyzing other information systems of the hierarchic type.

An Integral Model of an Information System of Hierarchic Type

Any system consists of elements and connections between them. The difference between individual systems is determined by the nature, number, and character of connections between the elements (they can be, for example, gravitational, nuclear, economic, or informational). Models of an atom in physics, structural formulas in chemistry, and models of economics are in essence attempts to penetrate the structure of systems of various types. Analogous problems are solved in linguistics by constructing structural models of a language, and also by developing information search languages and systems. For example, setting up of semantic codes, found in the information search system BIT developed under the leadership of É. F. Skorokhod'ko consists of constructing a tree-shaped graph which describes the structural connections between concepts and words of a text [12].

Specificality of individual systems does not exclude possibilities of constructing a model which would integrally describe certain properties that are common to all systems.

"Organizability," presence of information connections between elements, and a hierarchical structure are among such general properties. A hierarchical structure in turn predetermines the distribution of the elements of the system with respect to the levels.

The subsequent constructions have a meaning first and foremost for hierarchic systems with clearly expressed information connections between the elements. These connections, for example, can be texts, sets of scientific publication, or an information network which is obtained when forecasting on the basis of expert opinions.

We call the totality of properties which determine the place of an element at the appropriate level the "informativeness" of the element. Here we assume that the frequency of appearance of a basic attribute has a high correlation with informativeness, which is not formally definable. (Subsequently the justification of such an assumption will be shown.)

Such an essential attribute of appearance of informativeness in a system of periodic publication in any field of science is the number of articles on this theme in each journal, the number of references to these publications (authors), the frequency of encounter of individual words in a text of given length, and

Translated from Kibernetika No. 1, pp. 139-148, January-February, 1973. Original article submitted June 18, 1972.

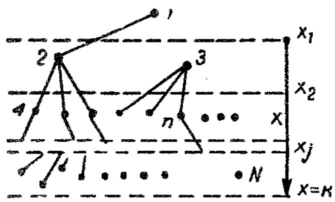

Fig. 1

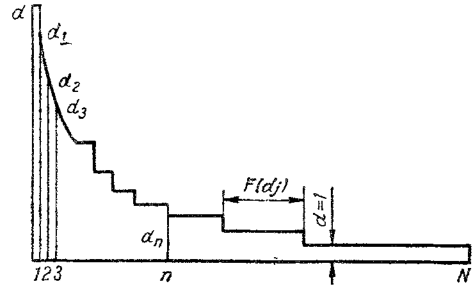

Fig. 2

similar items. Here, of course, we take into account the fact that the frequency of encounter of the attributes just enumerated or analogous attributes does not completely exhaust the informativeness of an element. A common property of hierarchic information systems is an exponential growth of the number of elements when we pass from an upper level to a lower level [13].

In Fig. 1 elements located on levels are numbered sequentially from 1 to N. Here $i(x=1) > i(x=2) > \ldots > i(x=K)$, where $i(x=K)$ is the informativeness of a single element of the k-th level.

It is obvious that a hierarchic system (Fig. 1) can be represented in the form of a distribution of a "hyperbolic" staircase (Fig. 2) for which the elements from 1 to N are ordered according to the frequency of the basic attribute (for example, according to the number of entries of a word from a dictionary into a text). With a view to generality when considering the hierarchical structure (Fig. 1) and its corresponding stepped distribution (Fig. 2), we introduce the following notation:

N is the number of different words of the text (a dictionary); the number of elements of the system corresponds to this;

n is the rank (number) of a word from the dictionary N, which is determinable by the number of entries of this word into the text (the distribution is ordered according to the ranks); informativeness of the element corresponds to the entry of the word into the text;

$d_1, d_2, \ldots, d_n$ is the absolute frequency of the word from the dictionary N respectively of the 1, 2, ..., n-th rank (the number of entries of the word into the text);

$p_1, p_2, \ldots, p_n$ is the relative frequency;

D_N is the overall number of words of the text with the dictionary N.

The meaning of the concepts d_n, p_n, D_N relative to the system of elements will be considered in detail later.

A step with the frequency of the basic attribute $(1, 2, \ldots, d_n, \ldots, d_1)$ will correspond to each level of the hierarchical system. If the number of elements increases exponentially when passing from level to level, then the depth of the system is proportional to the logarithm of the number (rank) of the element, i.e.,

$$X = x_0 + r \ln n. \tag{1}$$

Each of the levels of the system (or the steps of the distribution in Fig. 2) consists of $p_j(d=i)$ N elements, where $p_j(d=i)$ is the probability of the elements of N belonging to the j-th step (level) with a frequency d = i, i.e., $p_j(d=i) = \frac{N(d=i)}{N}$. Here $\sum_i p_j(d=i)N = N$, since $\sum_i p_j(d=i) = 1$.

Within a step (i.e., within the limits of a single level) the numbering of elements according to the ranks (numbers) is arbitrary, and therefore it is assumed that a change of the place of an element within the limits of ranks (numbers) which belong to the level (step) does not lead to a change in the state of the system. Within the limits of each level (step) there are $(p_jN)!$ possible permutations of the elements which do not alter the state of the system, while altogether in a hierarchical system of N elements there are $(p_1N)!\,(p_2N)!\ldots(p_kN)!$ such permutations.

"Organizability" of a system is defined as the logarithm of the ratio of possible number of permutations which do not change the system (i.e., permutations only within the limits of a single level) to the total number of permutations in the system of N elements, where all elements are assumed to be identical with respect to their functional possibilities (informativeness), i.e.,

$$I = -\ln \frac{(p_1N)!\,(p_2N)!\ldots(p_kN)!}{N!}. \tag{2}$$

According to the Stirling expression

$$\ln (p_kN)! \approx p_kN \ln p_kN - p_kN, \tag{3}$$

$$\ln N! \approx N \ln N - N. \tag{4}$$

Substituting (3) and (4) into (2) we obtain

$$I = -N \sum_k p_k \ln p_k. \tag{5}$$

The quantitative measure of organizability of a system, following E. Schrödinger [14] and L. Brillouin [15], is called negative entropy of the system which in this case characterizes "disorganizability" of the system. The problem consists of determining the conditions of maximum organizability of a hierarchical information system in the sense (2), (5). Formally this leads to the problem of variational calculus, of finding a distribution density (p) which minimizes the entropy integral

$$I = \int_0^\infty p(x) \ln p(x) dx \tag{6}$$

under the condition

$$\int_0^\infty p(x)\, dx = 1. \tag{7}$$

As the second condition, we take

$$M(X) = \int_0^\infty xp(x)\, dx = m. \tag{8}$$

The use for subsequent results of the given expected value M(X) and, consequently, the taking of the number of levels of the system in the role of the random quantity, are based on the possibility of interpreting the expected value as the coordinate of the "center of gravity" of the system. Here the points $X(x_1, x_2, \ldots, x_k)$ correspond to the depth of the hierarchical structure, to its levels, each of which in its turn is matched with a set of numbers of elements and the frequency of each of its basic attributes (according to which we measure the informativeness of the element). Thus, the value of the random quantity (X) is the number of elements of the level ($N_{X=j}$) and the frequency of attribute (informativeness) corresponding to the level.

Under the given conditions the search of p(x) which makes (5) a minimum is carried out by solving the following equation:

$$\frac{\partial F}{\partial p} + \lambda_1 \frac{\partial \varphi_1}{\partial p} + \lambda_2 \frac{\partial \varphi_2}{\partial p} + \ldots + \lambda_n \frac{\partial \varphi_n}{\partial p} = 0, \tag{9}$$

where

$$F = p \ln p; \quad \frac{\partial F}{\partial p} = \ln p + 1; \tag{9.1}$$

$$\varphi_1 = p; \quad \frac{\partial \varphi_1}{\partial p} = 1; \tag{9.2}$$

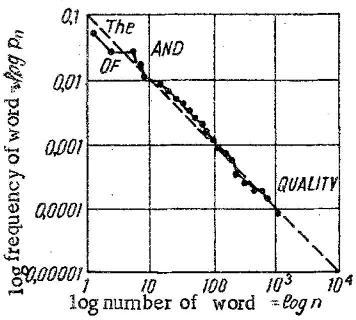

Fig. 3

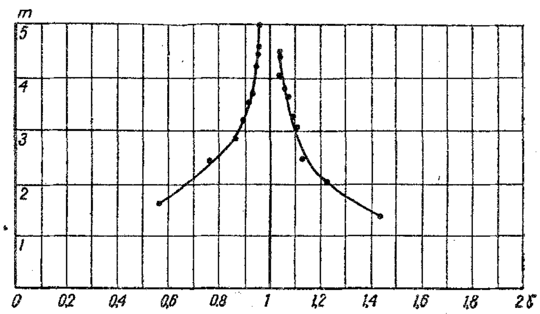

Fig. 4

$$\varphi_2 = xp; \quad \frac{\partial \varphi_2}{\partial p} = x. \tag{9.3}$$

After substitution of (9.1), (9.2), and (9.3) into (9), superposition λ_1 and λ_2 on the basis of the conditions (7), (8) and the necessary transformations, we obtain

$$p = \frac{1}{me^{\frac{x}{m}}}. \tag{10}$$

(For a detailed derivation of (10) see [11].)

In the expression (1) we determine the value of the quantities x_0 and r from the initial conditions and the requirement on p(x) and n(x) such that

$$p' \,|(x)| \sim n'(x), \quad n = \frac{1}{m} e^{\frac{x}{\gamma m}}, \tag{11}$$

where $r = \gamma m$. It is obvious that

$$\lim_{\gamma \to 1} \frac{p'|(x)|}{n'} = \frac{\frac{1}{m^2} e^{\frac{|x|}{m}}}{\frac{1}{m^2 \gamma} e^{\frac{x}{\gamma m}}} \to 1. \tag{12}$$

For x = 0

$$n = p = \frac{1}{m}; \tag{13}$$

for n = 1

$$x_0 = \gamma m \ln m. \tag{14}$$

Hence

$$X = \gamma m \ln n + \gamma m \ln m. \tag{15}$$

Substituting (15) into (10) and taking

$$\frac{1}{m^{\gamma+1}} = c, \tag{16}$$

we obtain

$$p = \frac{c}{n^{\gamma}}. \tag{17}$$

The expression (17) formally coincides with the well-known law of Zipf [6], who, on the basis of a statistical analysis, formulated this relationship. It was shown that (17) satisfactorily approximates the step function obtained by ordering the dictionary (N) according to the number of entries of each word of this dictionary into the text D_N. In Fig. 3 this relationship is illustrated in bilogarithmic coordinates [6].

In a number of investigations speculations were expressed that the optimizing properties of the system manifest themselves in the Zipf distribution: the minimum cost of an optimal code (B. Mandel'brot [7]), the maximum probability of a number of texts in the case of certain restrictions (Yu. A. Shreider [8]), the minimum value of the entropy of distribution (L. S. Kozachkov [10, 11]). The model being discussed here in a certain sense is a generalization of the speculations just enumerated.

The expression (17) (law of Zipf) contains the parameter γ whose value, by assertion of a number of authors having carried out a statistical analysis of a text, is given as close to unity [16, 17]. An estimate of the value of γ can be obtained from the analysis of the model itself. Under the condition that informativeness has a high correlation with the values of the essential attributes of the elements, the overall informativeness of the system of N elements ("lengths of the text") amounts to

$$D_N = \varphi(\gamma) = C \sum_{n=1}^{N} \frac{1}{n^\gamma} \approx \begin{cases} \frac{C}{1-\gamma} N^{1-\gamma}, \gamma < 1; & (18.1) \\ C \ln N, \gamma = 1; & (18.2) \\ \frac{C}{\gamma - 1}, \gamma > 1. & (18.3) \end{cases}$$

One of the important criteria of organizability of the system is the total maximum informativeness of a system in the case of a given distribution law $[d_n = (C/n^\gamma)]$ and a given number of elements (a dictionary) N. It is required to find such γ for which $D_N = \sum_{n=1}^{N} d_n$ has a maximum value. This value of γ is found by differentiating (18) with respect to γ and solving the equation for γ. For (18.1)

$$\frac{\partial \varphi}{\partial \gamma} = \frac{1}{1-\gamma} N^{1-\gamma} \left(\frac{1}{1-\gamma} - \ln N \right) = 0, \qquad \gamma = 1 - \frac{1}{\ln N}. \tag{19}$$

Formally D_N in (18.2) and (18.3) does not depend on γ.

From (19) we see that already for $N = 10^2$ $\gamma \approx 0.8$, while for $N = 10^5$–10^6 $\gamma \approx 0.9$. However, there are no real values of N for which γ becomes equal to unity (under the condition of an optimal organizability of the system). The value $N \to \infty$ has no physical meaning since the number of elements in any of the information systems considered is finite. The largest of known dictionaries contain hundreds of thousands of words (including synonyms and words which long ago have gone out of usage), systems of journals on certain topics contain hundreds of titles (there are altogether approximately $5 \cdot 10^3$ journals in the world), organizational structures contain up to a million elements, while a human brain contains approximately $1.4 \cdot 10^{10}$ neurons.

We shall next determine other parameters of the model of a hierarchical system. The parameter c by definition [see (16)] is $c = 1/m^{\gamma+1}$. However, it is not possible to obtain an estimate of c on the basis of (16), since the value of the "mean" depth M(X) = m of the system is not known. Therefore for the time being we determine c from the norming conditions

$$\sum_{n=1}^{N} p_n \approx \int_1^N p_n dn = 1 = \begin{cases} \frac{c}{1-\gamma} N^{1-\gamma}, 0 < \gamma < 1; & (20.1) \\ c \ln N, \gamma = 1; & (20.2) \\ \frac{c}{\gamma - 1}, \gamma > 1. & (20.3) \end{cases}$$

With (20) taken into account, we respectively have

$$c \approx \begin{cases} 1 - \gamma \text{ for } \gamma < 1; & (21.1) \\ \frac{1}{\ln N} \text{ for } \gamma = 1; & (21.2) \\ \gamma - 1 \text{ for } \gamma > 1. & (21.3) \end{cases}$$

C. Shannon assumed the value of the constant $c \approx 0.1$ for the determination of the "active" dictionary of texts in English language and the calculation of the entropy of a text [16]. The stability of the value c became the basis for the assertion by certain authors that c is the "universal" constant of a text [17]. "Universality" of this constant is more likely than anything connected with the informativeness of a "leader" (i.e., an element of the first rank of the distribution of words of the text).

Since $d_{min} = 1$, from $d_1 = \frac{C}{N^{\gamma}_{max}}$ we have

$$C \approx N^{\gamma}. \tag{22}$$

Up to now we considered $c = d_n/D_N$ and $C = cD_N$ in the role of the distribution constants and the model of the system. But such static treatment of the model of an information system is not obligatory. We can certainly assume that $C = \varphi(N)$; i.e., the informativeness of an element of the first rank can increase as N grows. In this case the static model becomes dynamic. Furthermore, if variation of γ in accordance with (19) and variation of C in accordance with (22) take place, then the system continues to preserve the optimal organization and maintain the maximum informativeness in the sense (18). If, however, $C = \text{const}$, then a growth of N leads to worsening of the organizability of the system. When passing from the static interpretation of the model to the dynamic interpretation, we can determine the conditions of achieving a maximum of informativeness in the case $\gamma > 1$ (18.3).

In the case $C = \text{const}$, D_N never reaches the maximum, since from $\partial D_N/\partial\gamma = C/(\gamma-1)^2 = 0$ it follows that D_N is maximum for $\gamma \to \infty$. The assumption that $C = N^{\gamma}$, in the case $\gamma > 1$, leads to the conclusion that for $D_N = N^{\gamma}/(\gamma-1)$ and $\partial D_N/\partial\gamma = N^{\gamma}\,[\ln N(\gamma-1)-1] = 0$

$$\gamma = 1 + \frac{1}{\ln N}. \tag{23}$$

Hence it follows that for $C = f(N)$

$$\gamma = 1 \pm \varepsilon, \tag{24}$$

where $\varepsilon = c \simeq 1/\ln N$ for $\gamma < 1$ and $\gamma > 1$.

These estimates of the value of γ agree with the investigations of a number of authors who have carried out a statistical analysis of a text and have discovered that cases $\gamma > 1$ take place, but the value of γ is found within the limits given by the expression (24). More accurate boundaries of variation of the parameter γ in a dynamic model will be determined below.

One of the central parameters of a model of a hierarchical structure is the "mean" depth $M(X) = m$ whose estimate is found from the conditions (21), (16):

$$M(X) = m = \begin{cases} \sqrt[\gamma+1]{\dfrac{1}{1-\gamma}} & \text{for } \gamma < 1; \\ \sqrt[\gamma+1]{\dfrac{1}{\gamma-1}} & \text{for } \gamma > 1. \end{cases} \tag{25}$$

Taking into account (19), for $\gamma < 1$ we have

$$m = \sqrt[1+\gamma]{\ln N} = (\ln N)^{\frac{\ln N}{2\ln N - 1}} \tag{26}$$

In Fig. 4 we have represented the graph $m = f(\gamma)$, which in more detail gives the boundaries of variation of γ:

$$\left.\begin{aligned} &\gamma \to 1, \; m \to \infty, \\ &\gamma \to 0, \; m \to 1 \end{aligned}\right\} \text{ for } \gamma < 1; \tag{27}$$

$$\left.\begin{aligned} &\gamma \to 2, \; m \to 1, \\ &\gamma \to 1, \; m \to \infty \end{aligned}\right\} \text{ for } \gamma > 1.$$

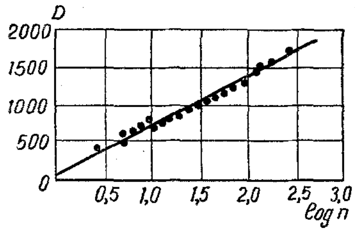

Fig. 5

TABLE 1

N	γ	m	c	K	$N_{x=2}$
10	0,566	1,7	0,434	1—2	~9
10^2	0,783	2,37	0,217	2—3	~6
10^3	0,856	3,0	0,144	5—6	~5
10^4	0,892	3,26	0,108	7—8	~5
10^5	0,913	3,6	0,087	10—11	~5
10^6	0,928	3,9	0,072	11—12	~5

An analysis of the functions (19), (25) and (26) shows that the values of γ having a physical meaning are bounded yet by the narrower intervals $0.95 \geq \gamma \geq 0.6$ ($\gamma < 1$) and $1.5 \geq \gamma \geq 1.05$ ($\gamma > 1$). Systems with a number of elements $N = 10\text{–}10^{10}$ correspond to this.

The expected number of levels of the system which ensures the maximum informativeness is found analogously to γ, if we consider $D_N = \varphi(X)$;

$$X = m(\ln N - 1). \tag{28}$$

From the expression (28) we determine the maximum number of elements of the second level, i.e., the estimate of a possible number of elements which directly are connected with the "leader" (an element of the first rank):

$$\ln N_{x=2} = \frac{2}{m} + 1. \tag{29}$$

The informativeness of a single level D_{x_j} is defined as $n_{x_j} \times d_{x_j}$, where n_{x_j} is the number of elements of the level X_j, and d_{x_j} is the informativeness of an element of the level X_j (for the example the case $\gamma < 1$ is considered). For $X_j = \gamma m \ln m$ (n = 1)

$$D_{x_j} = \frac{D_N}{m^{\gamma+1}}. \tag{30}$$

Under the condition that X can assume only integer values (1, 2, 3, . . ., K) when $\gamma \simeq 1$, the informativeness of the levels of the hierarchical information system will be the same and numerically equal to (30). Using (30), we can easily find the connection between M(X) = m and K, the total number of levels of the system:

$$\frac{D_N}{m^{\gamma+1}} = \frac{D_N}{K}; \tag{31}$$

$$m = \sqrt[\gamma+1]{K}. \tag{32}$$

However, with the fact taken into account that $/n'/ \sim /p'/$ only for $\gamma \simeq 1$, and with a correction introduced in view of this (essentially for the zone of small N), we finally have

$$K = \frac{m^{\gamma+1}}{e^{\frac{1-\gamma}{\gamma}}}. \tag{33}$$

It is not difficult to show that $dM(X)/dN \simeq 1/N\sqrt{\ln N}$ [the derivative of the function (26)]. It can be assumed in the estimate of real systems that, commencing with a certain N*, M(X) = m and the maximum number of levels of the system (X = K) becomes a "constant" and does not depend on N.

Generalizing the above analysis of the parameters of the model of a hierarchic information system, we write the general expression connecting all parameters of the system being considered:

$$m = \sqrt[\gamma+1]{K} = \sqrt[\gamma+1]{\ln N} = \sqrt[\gamma+1]{\frac{1}{c}} = \sqrt[\gamma+1]{\frac{1}{1-\gamma}} = (\ln N)^{\frac{\ln N}{2\ln N - 1}}. \tag{34}$$

In Table 1 we have represented the calculated values of γ, m, c, K, $N_{x=2}$ dependent on the number of elements of the hierarchical information system

The table is limited to the value $N = 10^6$. As was already mentioned, information systems with a large number of elements, although they exist, are less probable than in the interval $N = 10^2$–10^6. In addition, when using the continuous function (17) and all its consequences, we must take into account the fact that the basic parameters of the system – the number of elements and the informativeness of the element – have a discrete character (the minimum value of the frequency from which we estimate the informativeness of the element: one publication, one reference, one entry of a word into the text, and so forth). Hence, in accordance with (15) $x_{n=1} = \gamma m \ln m$ and $X_{max} = K$ we must not every time go to fractional values of the informativeness unit.

Certain Consequences and Possible Applications of the Model

Statistical Distributions in Informatics and Management of Science

1. Distribution (law) of Lotka: "The product of the frequency of scientists (Y_x) writing in a publication by the square of the number of publications (X) is a constant quantity" [3]:

$$Y_x X^2 = \text{const.} \tag{35}$$

2. Distribution (law) of Bradford: "If scientific journals are arranged in falling order of their productivity, i.e., with respect to the number of articles on a given problem, then they can be divided into basic periodic publications primarily devoted to the given problem, and several groups or zones containing the same amount of articles as the basic zone, and the number of publications in the basic group (core) and subsequent zones will be related as $1:n:n^2$" [4]:

$$N_1 : N_2 : N_3 : \ldots : N_k \approx 1 : \rho : \rho^2 : \ldots : \rho^{k-1}, \tag{36}$$

where N_k is the number of publications in the k-th zone, and ρ is a constant.

The relationships thus obtained were illustrated by Bradford on a field of publications on geophysics in the form of a straight line in the coordinates D_n, $\ln n$ (D_n is the cumulative number of publications; n is the number of a journal in an ordered set of publications, according to the number of publications on the given theme).

M. Kendall studied the given distribution using an example of a bibliography on an investigation into operations research and statistics and asserted that "linearity is first-rate" [5] (Fig. 5).

Hence it follows that

$$D_n = d_1 + \operatorname{tg} \varphi \ln n. \tag{37}$$

In the case where $\gamma \simeq 1$, from the Zipf distribution (17) we can easily obtain the Lotka and Bradford distributions. This is shown in a number of papers [5], [8], [9].

Indeed, (37) is the integral form of (17), while (35) is obtained from (17) by simple transformations.

The model of an information system described above allows us not only to discover the formal connection of these distributions with the Zipf distribution, but also enables us to penetrate into the general nature of these (and analogous) distributions which consist in the community of properties of any "organized" hierarchical information structure. In the case of such an approach the Zipf distribution itself, and other distributions connected with it, are particular cases of a hierarchical information structure, and values of the parameters characterizing the model of the hierarchical system occur in the analysis and use of these particular cases.

Systems of Scientific Publications

Side by side with factographic information, greater and greater importance is gained by problems of elucidation of the tendencies of development of individual fields [18]. In connection with this, information search systems (ISS) allowing us to analyze the field of information attract attention. Such an ISS whose field is a system of scientific publications is the Science Citation Index of U. Garfield. The field of this system consists of bibliographic description of articles (more than 1.5 thousand journals are scanned) and the references contained by them. In addition to the problem of literature search, this provides a possibility for carrying out a certain analysis of the field. A most striking example of the use of such networks was the study of interrelationships which lead to the deciphering of a heredity code [18].

"Informativeness" (number of references) of level	Numbers (ranks) of elements	Number of elements of level	Number of level
298	1–2	2	I
278	3–6	4	II
269	7–13	7	III
250	14–32	19	IV
291	33–173	141	V

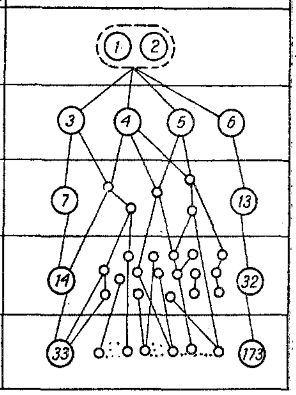

Fig. 6

A shortcoming of the SCI is the fact that only one type of information connection – references – is used, and that other languages developed in the practice of descriptor ISS are not used.

We have undertaken an attempt to develop a system which in principle increases the possibilities of analysis of the field, by broadening the language of connections [19]. The ASASI system (Automated System of Analysis of Scientific Information) was tested on a field of publications on the theory of automata and mathematical machines which was formed by the method of an index of scientific citations [20]. In Fig. 6 we have represented the block of a system of periodic publications.

In conformity with the expected theoretical value of the number of levels, the information "tree" of publications contains 5 levels.

Already 2 levels (6 journals of the first rank out of 173) contain more than 40% of informativeness of the entire system of publications. In the given case informativeness was determined by counting the number of references per publication and the number of articles published in them (the correlation between these two distributions $\rho \sim 0.8$ [20]).

The hierarchical system of periodic publications presented here can easily be shown in the form of a "hyperbolic staircase" distribution, and then it is described by the Bradford law (36). However, the law of "information dispersion" is only an approximate description of the fact of journal distribution over the information levels of the system. It is not a consequence, as is often assumed, of the chaos in the documentation science and publishing business. In addition, the features of the model of hierarchical information systems are such that it is more justifiable to talk about concentration and not dispersion of information. This conclusion is supported by many statistical investigations.

The language of the ASASI system is specifically developed for the analysis of fields constructed as a type of information network. In contrast to descriptor ISS, where the unit of the field is an individual document written in some language, the unit of the information field of the ASASI system is a generalized search image of a document (GSID) which includes both the basic document (primary source) and the references contained by it. At the same time the articles referred to by the author of the basic document are registered in the same language as the primary source. GSID consists of three parts:

1. Reference and bibliographic description. Here, in addition to the information that is traditional to a bibliography (name, first name, patronym, title of the document and similar items) special sections are included (for example, the type of work: survey, original, theoretical, experimental). The reference and bibliographic description is concluded by the construction of the search image of the title ("title descriptors").

2. Description of the grammatical structure (of all primary source and references). The facet analysis of the field thus carried out enabled the following lexical categories in the language to be singled out: 1) subject; 2) relations and properties; 3) object of investigation; 4) methods and algorithms; 5) results of investigation; 6) sphere of application.

This section of the language includes a selection of special indices of the role, and the horizontal and vertical connections (i.e., for determining the connections between the categories of a single document and between the categories of all documents that enter into the same GSID).

3. Abstract of 500 symbols. Below certain typical (sorting out) enquiries to the ASASI system are presented:

Formation of nonintersecting groups of documents with respect to the various criteria of semantic correspondence.

Construction of various distributions describing the frequency of a particular attribute or a collection of them, as well as the character of the connections between elements of the information field.

Construction of genealogical trees of some problem, when in the role of the starting problem a specified publication or a lexical category of the language (method, algorithm, subject, etc.) is chosen, and attributes of the basic table of the GSID language serve as constraints.

The following enquiry can serve as an example here: "Construct a genealogical tree of publications, taking Complex Systems and Solution of Extremal Problems as the initial work" [2]. The construction of the "tree" was constrained by two iterations: 1) references of the initial work; 2) references of the references. The following attributes, for example, can be constraints in the construction of the network: form of the source (monographs, articles); country (USSR, France, USA, Federal German Republic).

Other Hierarchical Information Structures

The methods of scientific and technical forecasting developed during the recent years are connected with the construction of graphs that are multilevel trees. This can be a tree of objectives (Pattern) or in the general case a network of scientific and technical events. According to the method of programmed forecasting and control of V. M. Glushkov, the experts structure the initial problem until the graph is earthed [21]. Available experience of programmed forecasting of a complex technical system showed that the first level of structuring of the initial problem had 7 classes; in each of these classes 5 to 10 alternatives of the solution of the problem were singled out, and each alternative included 7-10 scientific and technical conditions. The number of information levels in the case of forecasting is also bounded and, obviously, does not exceed 5-6, since already on the third level the experts put forward requirements on materials with specified properties, fundamental problems not amenable to structuring, or relatively simple earthing conditions. Although in the Pattern system [22] and SPI [23], the method of constructing the tree of objectives is different from the forecasting graph [21], here also the real forecasting networks had 5-7 levels.

Similar estimates of the number of levels and elements of each of the levels hold also for other classification systems, for example, for the UDC. The latest issues of the UDC contain more than 100 thousand headings [24]. Hence the expected number of levels must be equal to 10-11 (see the table of values for K). This estimate corresponds to the maximum depth of indexing in conformity with the "Tables of the UDC" [25]. Such a depth, however, is seldom used in practice. Calculation of the expected value of the depth of indexing for several journals of the cybernetics and computer technology gave an estimate of 5-7 signs. Of course, we are not dealing with multiaspect indexing, which in the framework of the UDC is difficult. However, in descriptor languages which are especially intended for multiaspect indexing, the search images also contain 8-10 descriptors, which is regarded as their optimal number [24], [26]. Search images with 3-5 descriptors are more often encountered, while those with more than 10 are seldom encountered.

Among the hierarchical information structures a special place is occupied by the natural language; the law of Zipf was in fact formulated on the basis of its statistical analysis. Being a "direct activity of thought," in the words of K. Marx, the language naturally reflects certain essential features of thinking. Precisely for this reason many properties of the language, including the Zipf law, are explained by the properties of thinking. For example, Zipf assumed that the distribution of words discovered by him is explained by the "principle of least effort" [6]. It is obvious that these optimizing functions of the brain must have influence on other information processes of memory and thinking. As was shown experimentally, a human being after a short period of observation can remember from 5 to 9 symbols [27]. Subsequently this phenomenon was several times repeated and refined. It was shown that these symbols can contain a different amount of information joined in blocks (words and entire phrases). Recording into the "operative" memory takes place in such a way that in the case of overfilling, its dispatch into the "long-term" memory takes place, but blocks of information are formed beforehand. With the functioning of the memory as a whole as a connected hierarchical system taken into account, the amount of information stored in the operative memory can be increased up to 10-12 symbols and even more. However, the expected value of the "capacity" of the operative memory remains within the limits 7 ± 2 symbols.

The maximum depth of phrases of the natural language discovered by Ingue has the same value [28]. It is assumed that a simple proposition has the depth of the tree of constituents equal to unity. The depth of complex trees of texts is bounded by the value $X \leq 9$, although cases of much deeper phrases are known, while in the German language phrases with the depth $X = 11$ are not a rarity [29].

The data about the mean and maximum numbers of morphemes of a word is also interesting — they also are found within the limits 5-9 [30] (in the language of our model this is the number of elements of the second level; see Table 1).

The astonishing stability of these figures in many information systems cannot be random, and must be explained by more general prerequisites. Therefore, as an explanation of the observed statistical properties of language and of the memory process (the depth of a phrase, the quantity of symbols in a single-level phrase, the number of symbols remembered in operational, i.e., short-term, memory, etc.), the boundedness of human memory is incomplete. The memory, being an organized hierarchical structure, has the same optimization features in the storage of information (and constraints) as other information systems (including the language which is being considered as an independent system), and not the other way round.

Although the agreement of the calculated parameters of the model with the experimental data is satisfactory, the use of the calculation expressions, when estimating real systems, requires caution. Firstly, complex information systems are not strictly deterministic but probabilistic; therefore large discrepancies can occur in real situations. Cases are known, for example, where authors "rose" against the requirements of the frequency of occurrence of individual letters and words. Thus one of the German poets in 130 poems not once used the letter R, while one of the English writers in a book 300 pages not once used the letter E. Moreover, there is no need to make a fetish, as is sometimes done, out of the number 7 ± 2. This estimate is only the expected value. Secondly, and this is the main point, the model proposed here merely integrally describes certain of the most general properties of information systems. For the analysis of the more refined properties far more refined models and methods of investigation are required. Thirdly, extension of the model beyond the limits of the experimental data on which it was formed (flows of scientific information and certain systems with clearly expressed information connections) is generally inadmissible. For example, the periodic table of chemical elements has 7 periods, while the maximum number of layers of an atom is also 7. But the elucidation of these facts is a task of chemistry and physics, and not that of information theory.

LITERATURE CITED

1. V. M. Glushkov, "On cybernetics as a science," in: Cybernetics, Thinking, Life [in Russian], Izd. Mysl', Moscow (1964).
2. V. S. Mikhalevich, Yu. M. Ermol'ev, V. V. Schkurba, and N. Z. Shor, "Complex systems and solution of extremal problems," Kibernetika, No. 5 (1967).
3. J. Lotka, "The frequency distribution of scientific productivity," J. Acad. Sci. (Washington), No. 12 (1926).
4. S. C. Bradford, Documentation, London (1948).
5. M. G. Kendall, "The bibliography of operational research," Operational Research Quarterly, Nos. 1, 2 (1960).
6. G. K. Zipf, Human Behavior and the Principle of Least Effort, Hafner (1965).
7. B. Madel'brot, "On recursive encoding limiting the influence of noise," in: The Theory of Communication Transmission [Russian translation], IL, Moscow (1957).
8. Yu. A. Shreider, "On the possibility of theoretical derivation of the statistical relationships of a text," Prob. Pered. Inform. No. 3, Moscow (1967).
9. L. S. Kozachkov, "On certain problems of relevance in informatics and management of science," Nauch.-Tekh. Inform., No. 8, Ser. 2, Moscow (1969).
10. L. S. Kozachkov, Investigation of Certain Systems of Documental Information Flows, Author's Abstract of Dissertation, Izd. VINITI, Moscow (1968).
11. L. S. Kozachkov, "Certain methodological problems of the theory of information search systems," Nauch.-Tekh. Inform., No. 12, Ser. 2, Moscow (1969).
12. "An information search system," in: BIT, A. A. Stognii (editor), Naukova Dumka, Kiev (1968).
13. S. Ya. Fitialov and G. S. Tseitin, "Estimates of the number of syntactic structures in the case of various constraints," Kibernetika, No. 6 (1966).
14. E. Schrödinger, What is Life?, Cambridge University Press.
15. L. Brillouin, Scientific Uncertainty and Information, Academic Press (1964).
16. C. Shannon, Works on the Theory of Information and Cybernetics [Russian translation], IL, Moscow (1963).
17. R. M. Frumkina, Statistical Methods of Study of Texts [in Russian], Nauka, Moscow (1946).
18. G. M. Dobrov, Science of Science [in Russian], Naukova Dumka, Kiev (1970).
19. L. S. Kozachkov, T. P. Dovgyallo, V. V. Moiseenko, A. A. Patiokha, and L. S. Potapenko, An Automated System of Analysis of Scientific Information (ASASI) [in Russian], Izd. IK Akad. Nauk USSR, Kiev (1972).

20. L. S. Kozachkov, "Certain problems of the theory of systems of information flows," Nauch.-Tekh. Inform., No. 10, Ser. 2, Moscow (1970).
21. V. M. Glushkov, "On forecasting on the basis of expert estimates," Kibernetika, No. 2 (1969).
22. E. Jantsch, Scientific-Technical Forecasting [Russian translation], Progress, Moscow (1970).
23. De L'Estoile, "La programmation de la reserche appliqué," Le Progress Scientifique, No. 118, Paris (1968).
24. A. I. Mikhailov, A. I. Chernyi, and R. S. Gilyarevskii, Foundations of Informatics [in Russian], Nauka, Moscow (1968).
25. Universal Decimal Classification [in Russian], Izd. GNTB, Moscow (1969).
26. F. W. Lancaster, Information Retrieval Systems, Wiley (1968).
27. J. A. Miller, "The magic number 7 ± 2. On certain limits of our capability to process information," in: Engineering Psychology, W. C. Howell and I. L. Goldstein (editors), Appleton (1971).
28. V. Ingve, "The depth hypothesis," in: New in Linguistics [in Russian], No. 4, Moscow (1965).
29. K. G. Pavlova, "Use of the method of V. Ingve in the determination of the depth of word combination," Uch. Zap., 61, Izd. Moskovskogo Pedinstituta Inostrannykh Yazykov (1971).
30. V. A. Moskovich, "Depth and length of words in natural languages," Vopr. Yazykoz., No. 7, Moscow (1967).