FTD-MT-24-414-69

# FOREIGN TECHNOLOGY DIVISION

## EXPERIENCE IN DEVELOPING, INTRODUCING AND USING DESCRIPTOR INFORMATION RETRIEVAL SYSTEMS UNDER SCIENTIFIC RESEARCH INSTITUTE CONDITIONS

by

Yu. I. Valov and A. V. Sokolov

# EDITED MACHINE TRANSLATION

EXPERIENCE IN DEVELOPING, INTRODUCING AND USING DESCRIPTOR INFORMATION RETRIEVAL SYSTEMS UNDER SCIENTIFIC RESEARCH INSTITUTE CONDITIONS

By: Yu. I. Valov and A. V. Sokolov

English pages: 17

Source: Nauchno-Tekhnicheskaya Informatsiya. Seriya 1. Organizatsiya i Metodika Informatsionnoy Raboty (Scientific and Technical Information. Series 1. Organization and Methods of Information work), 1968, No. 2, pp. 22-26.

This document is a Mark II machine aided translation, post-edited for technical accuracy by: Edwin P. Pentecost.

UR/0446-68-000-002

## U. S. BOARD ON GEOGRAPHIC NAMES TRANSLITERATION SYSTEM

| Block | Italic | Transliteration | Block | Italic | Transliteration |
|---|---|---|---|---|---|
| А а | *А а* | A, a | Р р | *Р р* | R, r |
| Б б | *Б б* | B, b | С с | *С с* | S, s |
| В в | *В в* | V, v | Т т | *Т т* | T, t |
| Г г | *Г г* | G, g | У у | *У у* | U, u |
| Д д | *Д д* | D, d | Ф ф | *Ф ф* | F, f |
| Е е | *Е е* | Ye, ye; E, e* | Х х | *Х х* | Kh, kh |
| Ж ж | *Ж ж* | Zh, zh | Ц ц | *Ц ц* | Ts, ts |
| З з | *З з* | Z, z | Ч ч | *Ч ч* | Ch, ch |
| И и | *И и* | I, i | Ш ш | *Ш ш* | Sh, sh |
| Й й | *Й й* | Y, y | Щ щ | *Щ щ* | Shch, shch |
| К к | *К к* | K, k | Ъ ъ | *Ъ ъ* | " |
| Л л | *Л л* | L, l | Ы ы | *Ы ы* | Y, y |
| М м | *М м* | M, m | Ь ь | *Ь ь* | ' |
| Н н | *Н н* | N, n | Э э | *Э э* | E, e |
| О о | *О о* | O, o | Ю ю | *Ю ю* | Yu, yu |
| П п | *П п* | P, p | Я я | *Я я* | Ya, ya |

* ye initially, after vowels, and after ъ, ь; e elsewhere. When written as ё in Russian, transliterate as yё or ё. The use of diacritical marks is preferred, but such marks may be omitted when expediency dictates.

# EXPERIENCE IN DEVELOPING, INTRODUCING AND USING DESCRIPTOR INFORMATION RETRIEVAL SYSTEMS UNDER SCIENTIFIC RESEARCH INSTITUTE CONDITIONS

Yu. I. Valov and A. V. Sokolov

For investigating the specific problems of creating and using descriptor systems in a local information organ by the information devision of NII [Scientific Research Institutes], is being conducted on the development and introduction into operation of a large descriptor IPS [Information Retrieval System] which can satisfy the basic mass of bibliographic demands placed upon the reference and information store [SIF] of NII. A system of descriptor IPS has been created which includes over 70 thousand documents and which services 70% of the requests made on the SIF. As the means of realization of the descriptor systems, peek-a-boo punched cards with a capacity of 1600 holes each are used. The initial assumptions and basic methodical solutions are described which are made in executing the work. Basic attention has been devoted to organizational problems of creating descriptor IPS under NII conditions. It is noted that in connection with the considerable tedium and complexity of works involved in the creation of descriptor systems, it is necessary to establish systematic and coordinated collaboration of different information services in this field.

The opinion regarding the need and prospect of developing and introducing descriptor IPS can today be considered universally accepted. However, the domestic practice of creating descriptor systems does not at present exceed the framework of limited experiments. The basis of

reference and information stores [SIF], as before, remain the main, manually retrieval information-bibliographic card files, systematized according to the Universal Decimal Classification [UDK] and formulated in accordance with traditional library-bibliographic rules. Therefore, any attempt to develop descriptor dictionaries of individual descriptor systems is of definite interest.

Of special value is an experiment of an information-bibliographic service based on the use of descriptor dictionaries or individual descriptor [Translator's Note: Russian text omitted here ] the results of the four year practice of development, introduction and exploitation of descriptor IPS in the information division of the NII.

The goal was set of studying the problem of creating in a local information organ, a large descriptor system as the basic means of an information-bibliographic service. Below methodical solutions are expounded; also, certain theoretical and organizational problems are posed which are connected with the expansion of propagation of descriptor IPS.

## 1. Initial Premises

It is obvious that the development of descriptor IPS is expedient only when, as a result of their introduction, an essential improvement in the reference and information service of consumers will be ensured. In order to meet this condition, it is important to know the deficiencies of the traditional SIF[1] information apparatus. A study of the spectrum of information requests proceeding to the ONTI [Dept. of Scientific and Technical Information] [1, 2], showed that the main drawback of traditional systematic and subject IPS is the great amount of losses of information and information noise. It is exactly this circumstance which involves the great tedium of retrieval according to narrow thematic requests which are connected with the industrial activity of engineering-technical workers. This means that the created descriptor systems must first be given the requirement of a more full and precise output of information than in traditional IPS.

A series of experiments conducted by us on a comparative investigation of the quality of work of traditional manual retrieval card files and descriptor systems [3-6] showed that a descriptor IPS can ensure very high retrieval quality, which, however, is not an inalienable property of descriptor systems, but is attained only under the condition of sufficiently great semantic force of the descriptor language when there is great depth and detail of indexing. The direction which must be adhered to during the development of descriptor IPS hence becomes clear: increasing the semantic force of the language and a comprehensive disclosure of the information contents of documents and requests in their retrieval forms.[2]

The above requirement on descriptor systems (ensuring an essential improvement of a reference and information service for consumers) was concretized in the form of defined requirements on component elements of IPS.

The experiments conducted, on the one hand, have convinced us of the possibility of creating high-quality descriptor IPS; on the other hand, then have shown the practical difficulties of their construction. It turned out that the development of IPS possessing great semantic force and ensuring deep and detailed indexing of documents requires considerable labor, especially in the first stage. The difficulties are heightened by the undeveloped nature of the theory of descriptor IPS and the low amount of practice in their usage at present. It follows from this that for creating descriptor systems it is necessary to have a sufficiently well qualified staff of designers.

Let us consider the concrete questions which appear in the work process.

## 2. Development of Micro-IPS

As already has been indicated, the basic problem of our work was to construct descriptor IPS capable of competing with the main information-bibliographic card file of an SIF. However, it is natural

to begin this work with the creation of a small descriptor IPS – a micro-IPS – in one of the thematic directions of the NII. On the basis of this micro-IPS methodical questions of the development of a descriptor dictionary practically can be checked and selection of optimum means of IPS realization can be carried out. Let us dwell on questions of the development of a descriptor dictionary for the micro-IPS (microthesaurus).

Inasmuch as we did not have any descriptor dictionaries in the Russian language (corresponding to the requirements set by us) the development of descriptor dictionaries was carried out by our own forces. We proceeded from the fact that the dictionaries created do not have to be combined terminological dictionaries on a given thematic topic, but should first of all ensure concrete retrieval functions, the primary of which is to ensure minimum losses and noise during information retrieval in a massif of documents introduced into a given IPS. Therefore, we used an inclusively empirical means of constructing the descriptor dictionaries, with which into the dictionary only that terminology was placed which corresponded to the information contents of the documents which were introduced into the IPS. Sources of terminology were only the key words which were selected from the text of primary sources or their annotations or abstracts. Considering the retrieval functions of the descriptor dictionary, we did not use any other terminology sources (classifications, linguistic and terminological dictionaries, textbooks, and others). Thanks to such a method of compilation, the descriptor dictionary of the micro-IPS took on orienting functions during retrieval, providing refusal to those requests which in no way were connected with the contents of the documents introduced into the micro-IPS. The empirical route of development is to us the simplest and most rational in the absence of other descriptor dictionaries on the given subjects.

It is obvious that the selection of key words directly determines the depth and detail of indexing in the descriptor language. In turn, on how qualitatively the descriptor retrieval form is composed finally depends the quality of retrieval in a given IPS. Unfortunately, at

present we do not have simple and clear recommendations concerning the indexing of key words; therefore, the selection of key words was made intuitively and was subordinated to the problem of achieving the most complete description of the contents of indexed documents. On the average, per document, depending upon the subjects of the micro-IPS, there were from 9 to 11 key words. The totality of selected key words was the glossary of the descriptor dictionary, which was subjected to further processing for removal of synonymy and for establishing the basic connections between descriptors.

Inasmuch as the theory of the descriptor systems (just as in the case of indexing key words) does not permit optimizing the composition process of descriptor dictionaries, a special experiment was conducted in which the rules of change of information losses and information noise losses were determined, depending upon the degree of depletion of the descriptor language and program of retrieval by request [7]. The experiment confirmed the presence of feedback between losses and noise, which consists of the fact that the introduction of means to combat the information noise inevitably leads to an increase of information losses, and vice versa. From this general theoretical assumption practical conclusions can be made. The use of means of calculating textual relations, which are intended to decrease information noise, has an effect on the increase of information losses. Grouping key words into classes of conditional equivalence (descriptors) is a means of decreasing losses, which involves an increase of information noise.

Inasmuch as in all cases a decrease of information losses is more desirable than a decrease of information noise, then in the development of the descriptor dictionary the following rule was accepted: the dictionary must have a maximum number of inputs into it, i.e., so that it considered the different terminological variations of the designation of ideas in the natural language, and at the same time so that it ensured reduction of the numerous inputs to a minimum number of outputs. In other words, the huge set of key words would be reduced to a small set of descriptors. A decrease of the number of descriptors, naturally, roughens the language and promotes the

appearance of information noise, but then high completeness of issue is ensured at this point.

Regarding programming the retrieval by request, the experiments showed a preference of such programs, in which the number of descriptors in the retrieval form of the request is minimum. For simplification of information retrieval in descriptor IPS, it is efficient to use the operation of excess indexing, the essence of which is that after formation of the "initial" retrieval form of the document, consisting of the descriptors caused by the context of the document, to the retrieval form descriptors are added which stand, according to the layout of the basic connections of the descriptor dictionary, higher than the initial ones; for example, in the presence of the descriptor "magazine" the descriptors "periodical," "literature," and "primary source of information" are added.

Of course, these considerations leave a great gap for subjective solutions in the development of a descriptor language, which remains a particularly creative process, depending first of all on the erudition and intuition of its designers. We await, from the theory of descriptor IPS, the answers to the following questions:

- formalization of indexing by key words;

- determination of the rules of presentation of word combinations and separate words in the form of descriptors:

- concretization of the concept of the "condition of equivalence" in reference to key words;

- formulation of the rules for exposure and fixing of basic connections between descriptors.

We accepted the following order of development of a descriptor dictionary for micro-IPS:

Stage 1. Indexing by key words of 800 documents – composition of the first editing of the descriptor dictionary of micro-IPS – a check of micro-IPS functioning under working conditions – dictionary correction.

Stage 2. Indexing by key words of the following 1600 documents – output of second editing of the micro-IPS descriptor dictionary (on a base of 2400 documents) – check of IPS functioning and dictionary correction.

Stage 3. Indexing by key words of the following 1600 documents – output of third editing of the micro-IPS descriptor dictionary (on a base of 4000 documents) – check of IPS functioning and dictionary correction, etc.

Implementation time of these operations was: stage 1 – 150 man days; stage 2 – 225 man days; stage 3 – 200 man days.

The most important question in the creation of a descriptor IPS is the selection of the means of realization. Attempting to ensure maximum simplicity and effectiveness of the system created, we refused to use an ETsVM [digital computer]. The remaining well-known means of realizing descriptor systems, namely, cards with edge performation, peek-a-boo punched cards and machine sorting punched cards were analyzed from the point of view of the following criteria: corresponds to the requirement of a minimization of losses and noise, convenience and speed of retrieval, and economic expenditure for realization and retrieval [8].

Correspondence to the first criterion was understood in the sense that the means of realization must not put any limitations on the correction and supplementation of the information-retrieval language, and they must also ensure easy variation of request retrieval routines. These conditions are best satisfied by peek-a-boo cards, which are not connected with the limited quality of a rigid mock-up. Minimum retrieval time in blocks of more than 10 thousand documents is ensured by peek-a-boo cards with a capacity of 7-10 thousand holes and higher,

which in this respect outstrip high speed electronic sorters, processing 1000 punched cards per minute. The expenditures of labor on input of information into IPS, realized on peek-a-boo cards, is somewhat higher than in the IPS which use cards with edge or slot punches; however, this deficiency is profitably compensated by lower labor expenditure of retrieval by request.

Proceeding from the conducted analysis, the conclusion was reached that the optimum means of realization of a descriptor IPS, intended for bibliographic retrieval under conditions of a local information organ, are peek-a-boo cards with capacity of 7-10 thousand holes.

Because these punched cards are not produced by domestic industry, it was decided to use, as the means of realization of our micro-IPS, standard 80-column punched cards which were used as the peek-a-boo cards. For increasing the capacity of these standard punched cards in the field between lines, additional holes were made, thanks to which the capacity of the punched card was increased to 1600 holes. This solution ensured ease of obtaining, storing and using the punched cards, since this does not necessitate special examination devices. For making holes with a 1.8 mm diameter a special punch and a mechanical pencil are used.[3]

Experience gained in the creation and use of several micro-IPS has confirmed that they can ensure a very high level of fullness and precision of issue (loss of information is 5-15%; information noise is 20-30%).

A description of the method of creating the micro-IPS and the equipment required is found in sources [9-12].

## 3. Transition from the Traditional SIF Information Apparatus to a System of Descriptor IPS

The creation of an individual narrow-thematic micro-IPS is not the solution to out problem. It is necessary to solve the problem of transition from a descriptor micro-IPS to a micro-IPS, which is

comparable by thematic width with branch library-bibliographic classifications. It is possible to proceed in two ways: expand the descriptor dictionary of the micro-IPS to the scale of a universal branch or general-scientific thesaurus, or try to create a system of interconnected micro-IPS which satisfies the majority of requests entering into a given SIF.

At the contemporary level of the theory of descriptor IPS it is possible to count on the creation of thesauruses of a universal character with a very small semantic force, equal in essence to the semantic force of universal library-bibliographic classifications of the Universal Decimal Classification type. Naturally, such thesauruses cannot ensure an essential improvement of the reference and information service, and, consequently, the success of measures undertaken for creating descriptor systems on their basis becomes doubtful. At the same time, the experience gained in development and use of micro-IPS has showed that for these systems a microthesaurus can be created with great semantic force, ensuring high completeness and precision of issue. Therefore, from the point of view of improving the quality of retrieval, a system formed by the totality of several micro-IPS is more prospective. But immediately the problem appears of delimitation of the subjects of micro-IPS which have been created.

With the contemporary integration and differentation of science and technology, the delimitation of separate thematic directions can be performed only in the case of acceptance of certain conventionalities. Several considerations say that it is rational to use the sum of conventionalities placed in the Universal Decimal Classification. In this case, to every micro-IPS a set of indices of the Universal Decimal Classification is placed in correspondence, forming the "record" of a given micro-IPS. According to this "record," into the given IPS only those documents are sent which carry one of the corresponding UDK indices. Thanks to such a solution, the transition from an information device, organized by the Universal Decimal Classification, to a system of a descriptor IPS is carried out without loss of the advantages of centralized reference and the information system, created at present on the basis of the Universal Decimal Classification; the succession

of both systems is ensured, and unity of reference and information stores on both the branch and country-wide scales is preserved. Also, interbranch information exchange and the system of centralized cataloguing are not disturbed.

The retrieval system constructed in such a manner can be called a classificational-descriptor IPS.[4] In this system the advantages of classificational systems (delimitation of thematic directions) and the descriptor systems (high quality of retrieval on the micro-IPS level) are combined.

The classificational-descriptor IPS possesses great flexibility, allowing the use of different methods of information retrieval. Retrieval by thematic divisions of the Universal Decimal Classification, which are not "profiling" for the given information organ, can be conducted manually, if expenditures for input of these divisions into the descriptor IPS appear groundless.

Thanks to the multiple aspect quality of retrieval in descriptor systems, in the micro-IPS documents different in form can easily be introduced: magazine articles, patents, reports on NIR [Scientific Research Work] and OKR [Experimental Design Work], etc. For books, monographs and collections concerning a great number of different subjects, an analytic list is obligatory which ensures identical "detail quality" of presentation of the documents introduced into the IPS.

As is noted above, the creation of a descriptor micro-IPS requires a great amount of labor. The greatest amount of labor was apportioned to the first stage, which is connected with the development of descriptor dictionaries. In this stage it is necessary to attract a great number of qualified specialists, possessing knowledge in the thematic region of the given micro-IPS and knowing the methods of constructing descriptor systems. It seems irrational to us to expand the state of an information service for carrying out essentially one-time operations on creating the bases of a classificational-descriptor system. It is simpler to attract to the work, conducted

by the information division, a group of engineering-technical workers of industrial subdivisions for consultations in the process of developing the descriptor dictionaries and for execution, when necessary, of technical work on formulating the micro-IPS.

After completing the first stage of creating classificational-descriptor IPS, all operations concerning service and augmentation of the system must be carried out by the efforts of information service colleagues. When planning the necessary labor involved it is possible to use the following experimentally-static norms, obtained by us for descriptor IPS carried out on peek-a-boo punched cards (during a 7-hour work day):

- indexing of key words and descriptors with a depth of indexation of 10 indices/document, 70 documents/day,

- auxiliary operations for preparation of mock-ups for punching the cards - 50 documents/day,

- card punch with realization of the operation of excess indexing and formulation of the card file - 40 documents/day.

Thus, the general cycle of input into the IPS of one document takes 25 minutes, from which only 1/4 of the time is spent on executing the work requiring engineering qualifications (indexing) and 3/4 of the time is spent on purely technical operations. Clearly, it is necessary to anticipate more expenditures of qualified labor on correcting the descriptor dictionary and on output of its subsequent editing. Expenditures on the output of the third, fourth, etc., editing of microthesauruses in our views, do not exceed 25 man days.

An important economic and organizational question in the development of descriptor micro-IPS is the realization of cooperation between the organizations interested in this work. Unfortunately, branch information institutes do not take effective measures for coordinating the efforts of local information organs in creating descriptor micro-IPS, and therefore collaboration between the separate

services is essentially an accidental affair. At the same time experience shows the possibility and rationality of exchange not only of descriptor dictionaries, but even of prepared micro-IPS. Thus, we transmitted to one of the adjacent (according to subject) NII two micro-IPS and, in turn, we obtained one micro-IPS corresponding to our subjects. Further augmentation of transmitted micro-IPS is carried out jointly. The planned division of labor between the separate organizations in the development of classification-descriptor systems seems to us to be a matter both realizable and necessary. Clearly, one should ensure single methodical principles of developing micro-IPS.

## 4. Service and Use of Descriptor IPS

In accordance with the above-stated principles, in our NII there has been created and now functions a classification-descriptor IPS, the characteristics of which are given in the table.

| Subjects of the micro-IPS | Indices of Universal Decimal Classification | Number of documents introduced into the IPS | Average number of descripters in the "initial" retrieval form | Number of editings of the descriptor dictionaries | Number of descripters/ number of key words (at last editing) |
|---|---|---|---|---|---|
| 1. Antennas and waveguides | 621.396.67 621.372.8 | 16 000 | 10,61 | 4 | 1183/1835 |
| 2. Transmitters | 621.396.61; 621.373 621.376 | 4 800 | 10,55 | 3 | 970/1310 |
| 3. Power supply | 621.311.6; 621.316.7; 621.314 | 5 800 | 9,42 | 3 | 403/558 |
| 4. Receivers | 621.396.62 | 1 800 | 10,75 | 1 | 699/913 |
| 5. Indicators | 621.396.963; 621.3a3 7/9; 621.396.832 621.317.755 | 6 400 | 10,35 | 3 | 761/1025 |
| 6. Amplifiers | 621.375 | 4 800 | 10,25 | 3 | 781/1063 |
| 7. Pulse technology | 621.374 | 5 600 | 9,35 | 3 | 637/880 |
| 8. Computer technology | 681.142 | 8 000 | 9,63 | 4 | 1197/1736 |
| 9. Reliability | 621.396.6.019.3 | 2 400 | 9,87 | 2 | 454/536 |
| 10. Microelectronics | 621.382.8 | 2 400 | 11,2 | 2 | 682/1300 |
| 11. Processing of signals | 621.391 | 3 000 | 8,09 | 2 | 1588/1848 |
| Totals: | | 59 900 | | 30 | |

As can be seen from the table, we found it possible to create a sufficiently large reference and information system, which ensures, according to data from the last half year, an answer to 70% of the

requests for bibliographic information in the SIP. The other 30% of requests are satisfied on the basis of the information-bibliographic apparatus, consisting of a system of traditional bibliographic card files and bibliographic publications. Experience has shown that a combination of the system of descriptor IPS with a traditional information apparatus is necessary, inasmuch as the subjects of a certain number of requests (nearly 10% of the overall amoung) cannot be anticipated when planning classificational-descriptor IPS, and, in connection with this, turning to traditional methods of bibliographic retrieval is inevitable.

Servicing the classificational-descriptor IPS is carried out by a group of bibliographers, a group of engineers-curators of micro-IPS (for these groups this function is not unique), and a group of technical workers. The bibliographers, knowing well the secondary sources of information, ensure making a set of micro-IPS in accordance with the "record" of the IPS in the Universal Decimal Classification language. They are also responsible for disseminating and collecting requests for bibliographic information. The obtained requests are sent to the corresponding micro-IPS or are satisfied by traditional bibliographic means.

Final selection of material for input into the micro-IPS is made by the engineer-curator of a given system. He compiles the "initial" retrieval form of the document in the descriptor language and makes, when necessary, a correction of the descriptor dictionary. Work on debugging the micro-IPS takes 30-40% of the engineer's time; the remaining time is designated for analytical and synthetic processing of information. The possibility of easy detection, systematization and correlation of information in the descriptor systems creates a good basis for increasing the quality of surveys, indicators and correlation tables, prepared by the curator.

Maintenance of the classificational-descriptor IPS is the responsibility of the group of technicians. According to our calculations, with a yearly entry into a classificational-descriptor IPS of 20 thousand documents, it will require 10 engineers spending 30-40%

of their work time with the IPS and 5 technicians, occupied with this work full time.

If when using the traditional means of information provision basic attention was allotted to servicing the upper-management echelons of the NII, then at present the possibility exists for more complete satisfaction of the information needs of ordinary engineers-designers. For the latter, of first priority value is concrete information about particular technical solutions, for which a high degree of completeness and accuracy of IPS issue is necessary. Practice has shown that descriptor systems considerably increase the effectiveness of a reference and information service.

The creation of descriptor systems has great value not only for improving informational work, but also for increasing the purposefulness of information concerning new entries. Experiments conducted by us showed that on the basis of the created descriptor languages a system of differentiated information distribution can be developed with the aid of digital computers.

## Conclusion

The principal result of our work is that it proved the practical possibility and expediency of developing and introducing effective information retrieval systems of the descriptor type under local information organ conditions. Along with the theoretical problems of developing descriptor languages in separate scientific and technical directions, organizational problems have been successfully solved which are connected with the order of carrying out works on mechanizing retrieval in SIF, with the delineation of functions between colleagues of an information service, with determining the order of augmentation and servicing of micro-IPS, etc. Presently available experience confirms the high efficiency of existing classificational-descriptor IPS, and the necessity of its further development gives no cause for doubts.

The question arises: is it possible to recommend that other local information organs proceed along our route? As far as we know, our experiment is now unique, which apparently, is not accidental. Completing the work we had planned was possible only thanks to a successful confluence of circumstances: the presence of trained cadres in the information division, on the one hand, and the possibility of using a large contingent of NII colleagues for carrying our one-time operations under the leadership of the information section, on the other. By far not in every NII and its information service are there such conditions. Obviously, it is irrational to proceed along the route of individual, disconnected, and essentially random "breakthroughs" in the field of descriptor systems. In order to solve complex theoretical and organizational problems connected with the development and introduction of a network of descriptor IPS, a systematic and directed "attack" is required.

The experience of our work, conducted outside the framework of centralized coordination, convinces one of the necessity for close collaboration of the information services of various classes.

## Bibliography

1. Sokolov A. V. Tipy informatsionnykh zaprosov spetsialistov i mekhamizatisiya poiska informatsii. (Types of information requests of specialists and mechanization of information retrieval.) "Tekhn bib-ki SSSR", 1966, No. 2, 16-22.

2. Venediktov Yu. P., Sokolov A. V. Izucheniye informatsionnykh zaprosov inzhenerov NII (A study of information requests from NII engineers). "ITI", 1964, No. 10, 7-9.

3. Sokolov A. V. Poteri informatsii i informatsionnyy shum v deskriptornykh IPS i v kartoteke, sistematizirovannoy po UDK (po rezul'tatam eksperimental'nogo issledovaniya) (Losses of information and information noise in descriptor IPS and in a card file, systematized by Universal Decimal Classification (according to the results of experimental investigation)). "NTI", 1964, No. 9, 20-23.

4. Sokolov A. V., Sorkin A. M. Stulencheskiy eksperiment v ovlasti nauchnoy informatsii (Student's experiment in the field of scientific information). "Biblioteki SSSR", 1966, No. 33, 163-165.

5. Sokolov A. V., Blyumenau D. I., Grinina R. F., Sorkin A. M. Eksperimental'nyye issledovaniya srabnitel'noy effektivnosti ruchnykh i mekhanizirovannykh IPS v LGIK im. N. K. Krupskoy (Experimental investigations of the comparative efficiency of manual and mechanized IPS in LGIK im. N. K. Krupskaya). Trudy Tret'yyey Vsesovuznoy konferentsii po IPS i avtomatizirobannoy obrabotke nauchno-tekhnicheskoy informatsii t. 1, M., 1967, 328-337.

6. Sokolov A. V. Istochniki poter' informatsii v traditsionnykh IPS i metody ikh kompensatsii (Sources of information losses in traditional IPS and methods of compensating them). "NII", 1967, No. 9.

7. Sokolov A. V. Issledovaniye poter' informatsii i informatsionnogo shuma v deskriptornykh IPS (Investigation of information losses and information noise in descriptor IPS). "NTI", 1965, No. 12, 23-23.

8. Sokolov A. V., Batashova L. Yu., Kozlova N. K. Mekhanizirovannyye informatsionno-poiskovyye sistemy mestnykh organov informatsii (Mechanized information retrieval systems of local information organs). 1965.

9. Sokolov A. V. Opyt mekhanizatsii informatsionnogo poiska v spravochno-informatsionnom fonde mestnogo organa informatsii (Experience in mechanization of information retrieval in the reference and information store of a local information organ). 1966.

10. Sokolov A. V. Superpozitsionnyye perfokarty kak sredstvo mekhanizatsii poiska v spravochno-informatsionnom fonde (Peek-a-boo cards as a means of retrieval mechanization in a reference and information store). In the book: "Sozdaniye informatsionno-poiskovykh sistem i informatsionno-poiskovykh yazykov dlya nikh". Chast' 2, 1965, 37-58.

11. Sokolov A. V. Obrabotka dokumentov dlya IPS, realizovannoy na superpozitsionnykh perfokartakh. Plakag serii "Mekhanizatsiya informatsionnogo poiska" (Processing documents for IPS realized on peek-a-boo cards. Placard of the series "Mechanization of information retrieval"). 1967.

12. Sokolov A. V. Poisk informatsii s pomoshch'yu superpozitsionnykh perfokart. Plakat serii "Mekhanizatsiya informatsionnogo poiska" (Information retrieval with the aid of peek-a-boo cards. Placard of the series "Mechanization of information retrieval"), 1967.

Article submitted
26 September 1967

## Footnotes

[1]By "traditional information apparatus" we understand a system of catalogs, card files and indicators of manual retrieval, organized by the systematic or subject principle with a degree of duplication of document descriptions of not more than 1.5 description/document.

[2]Experiments have shown that due to an increase of the depth of indexing and the degree of duplication of document descriptions in traditional IPS, the quality of retrieval in them can be improved; however, broader possibilities in this direction are found in descriptor systems which are based on the free coordination of single concepts

[3]Design of the pencil was developed at the NIIEIR [Translator's Note: Expansion unlisted].

[4]The idea of a classificational-descriptor IPS belongs to V. I. Vinogradov.

**DATA HANDLING PAGE**

| 01-ACCESSION NO. | 02-DOCUMENT LOC | 20-TOPIC TAGS |
|---|---|---|
| TP0500117 | | information storage and retrieval, information processing, descriptor dictionary, information descriptor system, punched card |

**03-TITLE** EXPERIENCE IN DEVELOPING, INTRODUCING AND USING DESCRIPTOR INFORMATION RETRIEVAL SYSTEMS UNDER SCIENTIFIC RESEARCH INSTITUTE *

**07-SUBJECT AREA**

09

| 08-AUTHOR CO-AUTHORS | 10-DATE OF INFO |
|---|---|
| VALOV, Yu. I. ; 16-SOKOLOV, A. V. | -----68 |
| **13-SOURCE** NAUCHNO-TEKHNICHESKAYA INFORMATSIYA. SERIYA 1. ORGANIZATSIYA I METODIKA INFORMATSIONNOY RABOTY (RUSSIAN) | **09-DOCUMENT NO.** FTD-MT-24-414-69 |
| | **09-PROJECT NO.** 6050205 |

| 63-SECURITY AND DOWNGRADING INFORMATION | 64-CONTROL MARKINGS | 97-HEADER CLASN |
|---|---|---|
| UNCL. 0 | NONE | UNCL |

| 76-REEL/FRAME NO. | 77-SUPERSEDES | 78-CHANGES | 40-GEOGRAPHICAL AREA | NO. OF PAGES |
|---|---|---|---|---|
| 1891 0568 | | | UR | 17 |
| **CONTRACT NO.** | **X REF ACC. NO.** | **PUBLISHING DATE** | **TYPE PRODUCT** | **REVISION FREQ** |
| | 45- | 94- | TRANSLATION | NONE |

**STEP NO.**
02-UR/0446/68/000/002/0022/0026


**ABSTRACT** * 09 CONDITIONS -U-

(U) For the purpose of studying the peculiarities of creation and operation of descriptor systems in a local information unit, the Information Department of a Research Institute is developing a large descriptor IRS capable of answering a bulk of requests received at the Institute's reference service. The system consists of subsystems, covering altogether 70,000 documents, and providing responses to 70 percent of requests. The system is implemented on peek-a-boo cards accomodating 1,600 holes each. The basic assumptions and decisions of policy are described. Creation of descriptor IRSs in the specific conditions of a research institute is in the focus of attention as an organizational problem. The authors points to the need for regular planned collaboration of various information services in creating descriptor systems which is a very involved and effort - consuming task.