AD 697 687

ALL-UNION CONFERENCE ON INFORMATION RETRIEVAL SYSTEMS AND AUTOMATIC PROCESSING OF SCIENTIFIC AND TECHNICAL INFORMATION (3RD), MOSCOW, 1967, TRANSACTIONS (SELECTED ARTICLES)

Foreign Technology Division
Wright-Patterson Air Force Base, Ohio

18 July 1969

FTD-MT-24-130-69

FOREIGN TECHNOLOGY DIVISION

ALL-UNION CONFERENCE ON INFORMATION RETRIEVAL SYSTEMS AND AUTOMATIC PROCESSING OF SCIENTIFIC AND TECHNICAL INFORMATION, 3rd, MOSCOW, 1967, TRANSACTIONS

(Selected Articles)

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce, for sale to the general public.

211

This document is a machine translation of Russian text which has been processed by the AN/GSQ-16(XW-2) Machine Translator, owned and operated by the United States Air Force. The machine output has been post-edited to correct for major ambiguities of meaning, words missing from the machine's dictionary, and words out of the context of meaning. The sentence word order has been partially rearranged for readability. The content of this translation does not indicate editorial accuracy, nor does it indicate USAF approval or disapproval of the material translated.

FTD-MT-24-130-69

EDITED MACHINE TRANSLATION

ALL-UNION CONFERENCE ON INFORMATION RETRIEVAL SYSTEMS AND AUTOMATIC PROCESSING OF SCIENTIFIC AND TECHNICAL INFORMATION, 3rd, MOSCOW, 1967, TRANSACTIONS (Selected Articles)

English pages: 187

SOURCE: Vsesoyuznoy Konferentsii po Informatsionno-Poiskovym Sistemam i Avtomatizirovannoy Obrabotke Nauchno-Tekhnicheskoy Informatsii, 3D, Moscow, 1967, Trudy, pp. 3-4, 9-15, 16-24, 25-29, 93-100, 134-142, 156-161, 168-171, 181-188, 201-208, 301-308, 328-337, 363-368, 454-476, and 477-482.

THIS TRANSLATION IS A RENDITION OF THE ORIGINAL FOREIGN TEXT WITHOUT ANY ANALYTICAL OR EDITORIAL COMMENT. STATEMENTS OR THEORIES ADVOCATED OR IMPLIED ARE THOSE OF THE SOURCE AND DO NOT NECESSARILY REFLECT THE POSITION OR OPINION OF THE FOREIGN TECHNOLOGY DIVISION.

PREPARED BY:

TRANSLATION DIVISION
FOREIGN TECHNOLOGY DIVISION
WP-AFB, OHIO.

FTD-MT-24-130-69 Date 18 July 1969

DATA HANDLING PAGE

01-ACCESSION NO. 98-DOCUMENT LOC TT9501081		39-TOPIC TAGS data processing conference, information storage and retrieval, machine translation, machine abstracting		
09-TITLE THE ROLE AND PLACE OF THE MACHINE IN SCIENTIFIC AND TECHNICAL INFORMATION				
47-SUBJECT AREA 05, 09				
42-AUTHOR/CO-AUTHORS MIKHAYLOV, A. I.				10-DATE OF INFO -----67
43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967 TRUDY (RUSSIAN)				68-DOCUMENT NO. FTD-MT-24-130-69 69-PROJECT NO. 6050205
63-SECURITY AND DOWNGRADING INFORMATION UNCL, O			64-CONTROL MARKINGS NONE	97-HEADER CLASH UNCL
76-REEL FRAME NO. 1889 1577	77-SUPERSEDES	78-CHANGES	40-GEOGRAPHICAL AREA UR	NO OF PAGES 11
CONTRACT NO.	X REF ACC. NO. 65-	PUBLISHING DATE 94-00	TYPE PRODUCT TRANSLATION	REVISION FREQ NONE
STEP NO. 02-UR/0000/67/001/000/0003/0004			ACCESSION NO.	

ABSTRACT

(U) This paper was presented at the All Union Conference on Information Retrieval Systems and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. The author depicts the place of the machine and its role in information activity and cites two independent, internally, interconnected problems: centralized processing scientific and technical literature and the creation of automatic input of printed text and satisfactory output from ETsVM of ready printed form for reproduction without typesetting, plus an entire group of complicated questions connected with the creation of information-retrieval systems. The need is stressed for the development of a wide front of theoretical and experimental works illuminating the problem of creation of information-retrieval systems.

DATA HANDLING PAGE

01-ACCESSION NO. 98-DOCUMENT LOC TT9501082	39-TOPIC TAGS data processing conference, data processing personnel, information storage and retrieval, computer language, information processing, computer technology, machine abstracting
09-TITLE BASIC TRENDS IN THE DEVELOPMENT OF INFORMATION RETRIEVAL SYSTEMS	
47-SUBJECT AREA 05, 09	

42-AUTHOR/CO-AUTHORS VOSKOBOYNIK, D. I. ; 16-VLEDUTS, G. E. ; 16-CHERNYAVSKIY, V. S.	10-DATE OF INFO -----67
43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)	68-DOCUMENT NO. FTD-MT-24-130-69 69-PROJECT NO. 6050205

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

76-REEL FRAME NO. 1889 1578	77-SUPERSEDES	78-CHANGES	40-GEOGRAPHICAL AREA UR	NO OF PAGES 14
CONTRACT NO.	X REF ACC. NO. 65-	PUBLISHING DATE 94-00	TYPE PRODUCT TRANSLATION	REVISION FREQ NONE

STEP NO. 02-UR/0000/67/001/000/0009/0015	ACCESSION NO.

ABSTRACT

(U) This paper was presented at the All Union Conference on Information Retrieval Systems and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. The author discusses the construction of a terminological machine dictionary, i.e. Thesauri, and the interrelationship between machine input and retrieval language, utilizing Thesaurus terms. Automatic indexing and reviewing are discussed on a general basis, with the author stating that the USSR has achieved favorable prerequisits for successful advancement in this area. The USSR appears to have many problems providing information Processing Systems with machine technology such as the comparatively low class of series domestic EVTsM; insufficient volumes of storage units, poorly developed parallelism of devices and low reliability, especially of external devices.

DATA HANDLING PAGE

01-ACCESSION NO. 98-DOCUMENT LOC TT9501083	39-TOPIC TAGS data processing conference, data processing personnel, information storage and retrieval, information processing, computer technology, intelligence training
09-TITLE PREPARATION OF ENGINEERING AND SCIENTIFIC CADRES WITH RESPECT TO MECHANIZATION AND AUTOMATION OF INFORMATION WORKS	
47-SUBJECT AREA 05, 09	

42-AUTHOR/CO-AUTHORS PETROV, I. I. ; 16-POCHKAY, I. B.	10-DATE OF INFO -----67
43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)	68-DOCUMENT NO. FTD MT-24-130-69 69-PROJECT NO. 6050205

63-SECURITY AND DOWNGRADING INFORMATION UNCL, O			64-CONTROL MARKINGS NONE	97-HEADER CLASN UNCL
76-REEL FRAME NO. 1889 1579	77-SUPERSEDES	78-CHANGES	40-GEOGRAPHICAL AREA UR	NO OF PAGES 7
CONTRACT NO.	X REF ACC. NO. 65-	PUBLISHING DATE 94-00	TYPE PRODUCT TRANSLATION	REVISION FREQ NONE
STEP NO. 02-UR/0000/67/001/000/0016/0024			ACCESSION NO.	

ABSTRACT

(U) This paper was presented at the All Union Conference on Information Retrieval Systems and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. The author states that up to 1965 there was no preparation of specialist to work in and further develop a system of scientific and technical information. In 1965 the Ministry of higher and special secondary education created in the higher school system, speciality No. 0640-automation and mechanization of processes of processing and delivery of information. Since the establishment of speciality No. 0640 improvement of the curricula has been developed, such as, disciplines of physicomathematical and engineering cycles must not contain obsolete information; a mathematics course must reflect the basis of mathematical logic, the principles of calculus of variation, probability and information theories, and other divisions needed to increase the mathematical level of the specialists in automation; the contents of disciplines in theoretical mechanics, strength of materials, theory of machines and mechanisms, descriptive geometry and drawing, and other general-engineering disciplines not directly related to specialities in automation should be radically examined, and the strengthening of general and special education of future automation engineers in electrical-engineering disciplines, electronics, and the general theory of automatic control.

DATA HANDLING PAGE

01-ACCESSION NO. 98-DOCUMENT LOC

TT9501084

39-TOPIC TAGS

data processing conference, data processing personnel, information storage and retrieval, machine abstracting, automatic document analysis

09-TITLE

THE LOGIC OF DESCRIPTOR RETRIEVAL SYSTEMS

47-SUBJECT AREA

05, 09

42-AUTHOR/CO-AUTHORS CHERNYAVSKIY, V. S.

10-DATE OF INFO -----67

43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)

68-DOCUMENT NO. FTD-MT-24-130-69

69-PROJECT NO. 6050205

63-SECURITY AND DOWNGRADING INFORMATION

UNCL, 0

64-CONTROL MARKINGS

NONE

97-HEADER CLASH

UNCL

76-REEL FRAME NO.	77-SUPERSEDES	78-CHANGES	40-GEOGRAPHICAL AREA	NO OF PAGES
1889 1580			UR	13
CONTRACT NO.	X REF ACC. NO.	PUBLISHING DATE	TYPE PRODUCT	REVISION FREQ
	65-	94-00	TRANSLATION	NONE

STEP NO.

02-UR/0000/67/001/000/0025/0029

ACCESSION NO.

ABSTRACT

(U) Two uniterm retrieval systems are analyzed- Pusto-Nepusto 2 and 4. The author states that comparison rules of the Pusto-Nepusto-2 system rests on different assumptions from the Pusto-Nepusto-4 system. These assumptions are formulated thusly: if an inquiry descriptor in a document is replaced with a lower descriptor this in no way reflects on the relevance of the document; if a document for a certain inquiry descriptor has not only an equal or lower but also a higher one, then this in no way effects document relevance; if in the document for a certain inquiry descriptor there is neither an equal nor a lower one, but at least one higher one, then this lowers document relevance, but does not make it equal to zero. Pusto-Nepusto-2 logic can split delivery into 2 echelons against the 4 echelons of the Pusto-Nepusto-4 system. The author concludes that the Pusto-Nepusto-2 system is in the experimental exploitation stage and it is too early to draw conclusions about the results of the conducted reorganization of the logic of the two systems.

DATA HANDLING PAGE

01-ACCESSION NO. **98-DOCUMENT LOC**

TT9501085

09-TITLE THE "SETKA-3" AUTOMATED IPS ON THE "MINSK-22" WITH THE USE OF THE SOCKET ASSOCIATIVE-ADDRESS METHOD OF ORGANIZATION OF INFORMATION

39-TOPIC TAGS

data processing conference, information storage and retrieval, information processing, computer technology, machine abstracting, automatic document analysis/(U)minsk digital computer

47-SUBJECT AREA

05, 09

42-AUTHOR/CO-AUTHORS GOROKHOV, S. A.

10-DATE OF INFO -----67

43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)

68-DOCUMENT NO. FTD-MT-24-130-69

69-PROJECT NO. 6050205

63-SECURITY AND DOWNGRADING INFORMATION	**64-CONTROL MARKINGS**	**97-HEADER CLASN**
UNCL, O	NONE	UNCL

76-REEL FRAME NO.	**77-SUPERSEDES**	**78-CHANGES**	**40-GEOGRAPHICAL AREA**	**NO OF PAGES**
1889 1581			UR	15
CONTRACT NO.	**X REF ACC. NO.**	**PUBLISHING DATE**	**TYPE PRODUCT**	**REVISION FREQ**
	65-	94-00	TRANSLATION	NONE

STEP NO. 02-UR/0000/67/001/000/0093/0100

ACCESSION NO.

ABSTRACT

(U) This paper was presented at the All Union Conference on Information Retrieval Systems and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. The author discusses the construction of automated information-retrieval systems of the descriptor type on the Minsk 22 and the Minsk 2 digital computer. The paper examines an improved variant of the Setka-3 information processing system on the Minsk 22 with use of the socket associative-address method of organization of information. Also examined in the paper are the creation of a dictionary of descriptors for the Thematic Division-Computer Technology; construction of a machine dictionary of descriptors of an automated IPS; processing of inquiries and their input into ETsVM; recording of initial information in ETsVM; and a description of the algorithm of work of the IPS.

DATA HANDLING PAGE

01-ACCESSION NO. 98-DOCUMENT LOC TT9501086	39-TOPIC TAGS data processing conference, information storage and retrieval, information processing, computer technology, computer language, machine abstracting, automatic document analysis
09-TITLE EXPERIENCE IN CREATING INFORMATION-RETRIEVAL LANGUAGE VIA COMPUTER TECHNOLOGY	
47-SUBJECT AREA 05, 09	

42-AUTHOR/CO-AUTHORS VAKHABOV, V. K. ; 16-MIKHAYLOVA, A. A. ;16-YESILEVSKAYA, L. M. ; 16-KUTAYEVA, T. S.	10-DATE OF INFO -----67
43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)	68-DOCUMENT NO. FTD-MT-24-130-69 69-PROJECT NO. 6050205

63-SECURITY AND DOWNGRADING INFORMATION UNCL, 0			64-CONTROL MARKINGS NONE	97-HEADER CLASN UNCL
76-REEL FRAME NO. 1889 1582	77-SUPERSEDES	78-CHANGES	40-GEOGRAPHICAL AREA UR	NO OF PAGES 9
CONTRACT NO.	X REF ACC. NO. 65-	PUBLISHING DATE 94-00	TYPE PRODUCT TRANSLATION	REVISION FREQ NONE
STEP NO. 02-UR/0000/67/001/000/0134/0142			ACCESSION NO.	

ABSTRACT

(U) This paper was presented at the All Union Conference on Information Retrieval Systems and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. The author reports on the development of information retrieval language of the descriptor type according to the division of computer technology. Conclusions of the article indicate that the developed information retrieval language (IPYa) has satisfactory characteristics; growth of a dictionary is considerably delayed during the growth of an array of over 2000 documents; input of grammatical means into IPYa is inexpedient for small arrays; the question regarding the need to introduce grammatical means into developed IPYa will be examined after conducting experiments on 15-20 thousand documents in 1967; and distribution of descriptors in retrieval patterns of documents obeys the same Zipf and Mandel'brot laws, as words in natural language, texts.

DATA HANDLING PAGE

01-ACCESSION NO. 98-DOCUMENT LOC TT9501087	39-TOPIC TAGS data processing conference, information storage and retrieval, information processing, computer technology
09-TITLE REALIZATION OF THEMATIC INFORMATION-RETRIEVAL SYSTEMS ON AN ELECTRONIC DIGITAL COMPUTER	
47-SUBJECT AREA 05, 09	

42-AUTHOR/CO-AUTHORS UBIYKO, M. P.	10-DATE OF INFO -----67
43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)	68-DOCUMENT NO. FTD-MT-24-130-69 69-PROJECT NO. 6050205

63-SECURITY AND DOWNGRADING INFORMATION UNCL, O	64-CONTROL MARKINGS NONE	97-HEADER CLASN UNCL

76-REEL FRAME NO. 1889 1583	77-SUPERSEDES	78-CHANGES	40-GEOGRAPHICAL AREA UR	NO OF PAGES 7
CONTRACT NO.	X REF ACC. NO. 65-	PUBLISHING DATE 94-00	TYPE PRODUCT TRANSLATION	REVISION FREQ NONE

STEP NO. 02-UR/0000/67/001/000/0156/0161	ACCESSION NO.

ABSTRACT

(U) This paper was presented at the All Union Conference on Information Retrieval Systems and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. The paper discusses the need to employ computer technology for information storage and retrieval of patents.

DATA HANDLING PAGE

01-ACCESSION NO. 98-DOCUMENT LOC

TT9501088

39-TOPIC TAGS

data processing conference, information storage and retrieval, automatic document analysis, machine abstracting, information processing

09-TITLE CERTAIN QUESTIONS OF MACHINE PREPARATION OF INDEXES FOR CURRENT AND RETROSPECTIVE RETRIEVAL OF SCIENTIFIC AND TECHNICAL *

47-SUBJECT AREA

05, 09

42-AUTHOR CO-AUTHORS MEKHTIYEV, D. I. ; 16-KUZNETSOVA, E. K.

10-DATE OF INFO -----67

43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)

68-DOCUMENT NO. FTD-MT-24-130-69

69-PROJECT NO. 6050205

63-SECURITY AND DOWNGRADING INFORMATION UNCL, O

64-CONTROL MARKINGS NONE

97-HEADER CLASN UNCL

76-REEL FRAME NO. 1889 1584	77-SUPERSEDES	78-CHANGES	40-GEOGRAPHICAL AREA UR	NO OF PAGES 12
CONTRACT NO.	X REF ACC. NO. 65-	PUBLISHING DATE 94-00	TYPE PRODUCT TRANSLATION	REVISION FREQ NONE

STEP NO. 02-UR/0000/67/001/000/0168/0171

ACCESSION NO.

ABSTRACT * 09 INFORMATION MATERIALS

(U) This paper was presented at the All Union Conference on Information Retrieval Systems and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. This paper discusses automatic indexing based on keywords. Recordings were processed on the Ural-4 according to a specially composed program: reproduction of titles according to the number of noted key words, intramachine sorting of reproduced titles according to the alphabet of key words in the input column of the UKS and printing. The author states that the creation and wide use of permutation indexes will allow operationally in more compressed periods informing consumers of information about the latest publications, increasing the effectiveness of current bibliography and making it accessible to all categories of specialists.

DATA HANDLING PAGE

01-ACCESSION NO. TT9501089

98-DOCUMENT LOC

39-TOPIC TAGS data processing conference, information storage and retrieval, automatic document analysis, machine abstracting, information processing, punched card

09-TITLE MECHANIZED IPS FOR A VARIED FUND DEVELOPED AND INTRODUCED BY THE UPPER-VOLGA TsBTI

47-SUBJECT AREA 05, 09

42-AUTHOR/CO-AUTHORS MESHIK, V. M.

10-DATE OF INFO -----67

43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)

68-DOCUMENT NO. FTD-MT-24-130-69

69-PROJECT NO. 6050205

63-SECURITY AND DOWNGRADING INFORMATION UNCL, O

64-CONTROL MARKINGS NONE

97-HEADER CLASN UNCL

76-REEL FRAME NO. 1889 1585

77-SUPERSEDES

78-CHANGES

40-GEOGRAPHICAL AREA UR

NO OF PAGES 15

CONTRACT NO.

X REF ACC. NO. 65-

PUBLISHING DATE 94-00

TYPE PRODUCT TRANSLATION

REVISION FREQ NONE

STEP NO. 02-UR/0000/67/001/000/0181/0188

ACCESSION NO.

ABSTRACT

(U) This paper was presented at the all Union Conference on Information Retrieval Systems and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. The author stated that the Information Processing System developed by the USSR is applied in practical activity of varied SIF during two years. Mechanized retrieval and delivery of information with the use of comparatively cheap punch-card equipment considerably reduces the time necessary for retrieval of documents and facilitates the labor of workers of the section. Although retrieval is conducted according to a limited number of retrieval criteria (4-5) in the majority of cases the USSR is satisfied with the results.

DATA HANDLING PAGE

01-ACCESSION NO. 98-DOCUMENT LOC

TT9501090

39-TOPIC TAGS

data processing conference, information storage and retrieval, information processing, computer technology

09-TITLE QUESTIONS OF DEVELOPMENT OF AN INFORMATION-RETRIEVAL SYSTEM FOR AUTOMATED CONTROL OF THE WORK OF A NAVAL FLEET OF STEAM NAVIGATION

47-SUBJECT AREA

05, 09

42-AUTHOR/CO-AUTHORS KISELEV, A. N. ; 16-MIKHAYLOVA, I. A.

10-DATE OF INFO -----67

43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967 TRUDY (RUSSIAN)

68-DOCUMENT NO. FTD-MT-24-130-69

69-PROJECT NO. 6050205

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

76-REEL FRAME NO. 1889 1586	77-SUPERSEDES	78-CHANGES	40-GEOGRAPHICAL AREA UR	NO OF PAGES 14
CONTRACT NO.	X REF ACC. NO. 65-	PUBLISHING DATE 94-00	TYPE PRODUCT TRANSLATION	REVISION FREQ NONE

STEP NO. 02-UR/0000/67/001/000/0201/0208

ACCESSION NO.

ABSTRACT

(U) This paper was presented at the all Union Conference on Information Retrieval System and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. The authors discussed improvement of the control system of a naval fleet of steam navigation by mathematics and computer technology. The IPS for automated control of the work of the fleet was developed in the following sequence: analysis of the problems of control, information communications; foundation of SAU in the form of an IPS; classification of objects and their characteristics; development of methods of distribution of information in the memory unit; development of algorithms and programs of input and restoration of information; and the development of algorithms and programs of answers to inquiries.

DATA HANDLING PAGE

01-ACCESSION NO. 90-DOCUMENT LOC

TT9501091

39-TOPIC TAGS

data processing conference, data processing personnel, information storage and retrieval, information processing, computer technology

09-TITLE EXPERIMENTAL INVESTIGATIONS OF COMPARATIVE EFFECTIVENESS OF MANUAL AND MECHANIZED IPS IN THE N. K. KRUPSKAYA LENINGRAD STATE *

47-SUBJECT AREA

05, 09

42-AUTHOR/CO-AUTHORS SOKOLOV, A. V. ; 16-BYUMENAU, D. I.; 16-GRININA, R. F. ; 16-SORKIN, A. M.

10-DATE OF INFO

-----67

43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)

68-DOCUMENT NO.

FTD-MT-24-130-69

69-PROJECT NO.

6050205

63-SECURITY AND DOWNGRADING INFORMATION	64-CONTROL MARKINGS	97-HEADER CLASH
UNCL, 0	NONE	UNCL

76-REEL FRAME NO.	77-SUPERSEDES	78-CHANGES	40-GEOGRAPHICAL AREA	NO OF PAGES
1889 1587			UR	15

CONTRACT NO.	X REF ACC. NO.	PUBLISHING DATE	TYPE PRODUCT	REVISION FREQ
	65-	94-00	TRANSLATION	NONE

STEP NO.	ACCESSION NO.
02-UR/0000/67/001/000/0301/0308	

ABSTRACT * 09 INSTITUTE OF CULTURE

(U) This paper was presented at the All Union Conference on Information Retrieval Systems and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. The paper discusses the experimental investigation of manual and mechanized information processing system in the N. K. Krupskaya Leningrad State Institute of Culture. Results of the investigation indicated that it was practically impossible to construct bibliographic IPS possessing an ideal quality of work, i.e., zero information losses, traditional IPS have approximately 40 percent higher information loss than model descriptor IPS, and descriptor IPS must provide a minimum level of information losses.

DATA HANDLING PAGE

01-ACCESSION NO. TT9501092

98-DOCUMENT LOC

39-TOPIC TAGS data processing conference, information storage and retrieval, information processing, computer technology

09-TITLE INFORMATION RETRIEVAL SYSTEMS

47-SUBJECT AREA 05, 09

42-AUTHOR CO-AUTHORS TRUKHAYEV, R. I. ; 16-KHOMENYUK, V. V.

10-DATE OF INFO -----67

43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)

68-DOCUMENT NO. FTD-MT-24-130-69

69-PROJECT NO. 6050205

63-SECURITY AND DOWNGRADING INFORMATION UNCL, O

64-CONTROL MARKINGS NONE

97-HEADER CLASN UNCL

76-REEL FRAME NO. 1889 1588

77-SUPERSEDES

78-CHANGES

40-GEOGRAPHICAL AREA UR

NO OF PAGES 10

CONTRACT NO.

X REF ACC. NO. 65-

PUBLISHING DATE 94-00

TYPE PRODUCT TRANSLATION

REVISION FREQ NONE

STEP NO. 02-UR/0000/67/001/000/0328/0337

ACCESSION NO.

ABSTRACT

(U) This paper was presented at the All Union Conference on Information Retrieval Systems and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. The authors describe an information retrieval system, its functional stages and structure.

DATA HANDLING PAGE

01-ACCESSION NO. 98-DOCUMENT LOC

TT9501093

09-TITLE A SYSTEM OF AUTOMATIC DIFFERENTIATION OF DISTRIBUTION OF INFORMATION (SARI-1) ON CONSTANT INQUIRIES DEVELOPED IN TsBNTI AKIAE[1]

47-SUBJECT AREA

05, 09

39-TOPIC TAGS

data processing conference, information storage and retrieval, information processing, computer technology

42-AUTHOR CO-AUTHORS NADTOCHIY, A. I. ; 16-KALININ, V. F. ; 16-MIKHEYEV, N. N. ; 16-VORONIN, V. A. ; 16-GOSTEV, V. I.

10-DATE OF INFO -----67

43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)

68-DOCUMENT NO. FTD-MT-24-130-69

69-PROJECT NO. 6050205

63-SECURITY AND DOWNGRADING INFORMATION	64-CONTROL MARKINGS	97-HEADER CLASN
UNCL, O	NONE	UNCL

76-REEL FRAME NO.	77-SUPERSEDES	78 CHANGES	40-GEOGRAPHICAL AREA	NO OF PAGES
1889 1589			UR	36
CONTRACT NO.	X REF ACC. NO.	PUBLISHING DATE	TYPE PRODUCT	REVISION FREQ
	65-	94-00	TRANSLATION	NONE

STEP NO.	ACCESSION NO.
02-UR/0000/67/001/000/0363/0368	

ABSTRACT

(U) This paper was presented at the All Union Conference on Information Retrieval Systems and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. In order to s lve the problem of information services satisfactory both in the time element and subject an automatic differentiation distribution of information with respect to constant inquiries was developed with the help of the Minsk-22 and based on a descriptor language. This paper discusses this automatic system.

DATA HANDLING PAGE

01-ACCESSION NO. 98-DOCUMENT LOC TT9501094	39-TOPIC TAGS data processing conference, information storage and retrieval, information processing, computer technology
09-TITLE APPLICATION OF DOMESTIC PUNCH-CARD MACHINES AND THE "MINSK-22" FOR COMPOSITION OF A DESCRIPTOR INDEX	
47-SUBJECT AREA 05, 09	

42-AUTHOR CO-AUTHORS NADTOCHIY, A. I. ; 16-KRASOVSKAYA, L. N. ; 16-MARTYNOV, L. P. ; 16-MIKHEYEV, N. N.	10-DATE OF INFO -----67
43-SOURCE VSESOYUZNOY KONFERENTSII PO INFORMATSIONNO-POISKOVYM SISTEMAM I AVTOMATIZIROVANNOY OBRABOTKE NAUCHNO-TEKHNICHESKOY INFORMATSII 3D, MOSCOW, 1967, TRUDY (RUSSIAN)	68-DOCUMENT NO. FTD-MT-24-130-69
	69-PROJECT NO 6050205

63-SECURITY AND DOWNGRADING INFORMATION UNCL, 0			64-CONTROL MARKINGS NONE	97-HEADER CLASN UNCL
76-REEL FRAME NO. 1889 1590	77-SUPERSEDES	78-CHANGES	40-GEOGRAPHICAL AREA UR	NO OF PAGES 9
CONTRACT NO.	X REF ACC. NO. 65-	PUBLISHING DATE 94-00	TYPE PRODUCT TRANSLATION	REVISION FREQ NONE
STEP NO. 02-UR/0000/67/001/000/0454/0476			ACCESSION NO.	

ABSTRACT

(U) This paper was presented at the All Union Conference on Information Retrieval Systems and Automatic Processing of Scientific and Technical Information held in Moscow in 1967. The paper discusses document distribution to users at TsBNTI utilizing a descriptor index. The index is intended both for users working directly in the field of atomic science and technology and users of other departments for which information of a similar form is needed. Output of the descriptor indices is one of the steps in the creation of an effective system of information services. Existing computerized IPS in the USSR at present can service only a certain circle of users governed by the possibilities of service of information and economic considerations. The authors indicate that the descriptor index should give wider distribution of documents utilizing the existing data processing system.

TABLE OF CONTENTS

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.

PREFACE

The Party and Government are giving their constant attention to the development and improvement of systems of the information service. The special resolution adopted by the Council of Ministers of the USSR in November 1966 is directed towards implementing solutions to the problems enacted by the XXIII Congress of the CPSU on creation of government-wide highly effective scientific-technical information service.

Being based on contemporary technical means for acquisition, processing, investigation and delivery of information data, and automation of information processes, the State system of scientific information is directed to provide timely information about achievements in domestic and foreign science and technology.

At present in many organizations of the country scientific research work is being conducted in the area of automation of information processes. The level and state of this work require defined organization and coordination. This is especially important since scientific and technical information is such a many-faceted branch of science and technology, embracing all branches of the national economy, that its further effective development would be inconceivable without coordination of the work and exchange of the experience of all specialists occupied with automation of information processes.

The Third All-Union Conference on information retrieval systems and automated processing of scientific and technical information, conducted from 19 through 22 December 1966 in Moscow, provided results of scientific and technical information activity of scientists and specialists and outlined the way for the fullest and fastest fulfillment of problems set by the Party and Government in the area of scientific and technical information.

The work of conference involved the participation of 1150 specialists representing different ministries, departments, information organs, scientific research and design organizations, industrial enterprises and establishments of the country. At the conference 220 reports were presented.

The present Transactions contain material presented at the plenary sessions of conference, as well as at meetings of the separate sections.

For the convenience of readers, the Transactions are being issued in four volumes.

The first volume "Information Retrieval Systems" presents the results of research and development in the areas: semantic systems of investigation of scientific and technical literature in large data bases, with automatic translation from a natural into a formalized language; automated systems of factographic facilities based on the use of information languages of the natural sciences; information retrieval systems for industrial enterprises and establishments.

The same volume is devoted to other questions connected with the creation and introduction of automated information retrieval systems of different classes and assignment for processing large as well as small volumes of scientific and technical information.

The second volume, "Semiotic Problems of Automated Data Processing" presents material dedicated to: development of problems of the connection between syntactic and semantic properties of language systems; investigation of the natural and formalized languages of science and technology in connection with problems of storage and retrieval of information; questions of automatic processing of tests for creating operational systems of machine indexing, abstracting and translation of texts; research in the area of creation of special programming languages and translators from them for machine processing of texts.

The third volume, "Automatic Reading Devices" presents results of research and development of different methods and devices for automatic identification of typographical and typewritten symbols. Special attention is alloted here to automatic reading machines, allowing automated computer input of masses of scientific and technical, statistical and economic information to ETsVM [electronic digital computers].

The fourth volume, "Technical Devices for the Information Service and on Line Reproduction Techniques" presents works related to research and development of: technical devices for preparation and input of alpha-numeric information into a computer and high speed output devices of textual information, and also output devices with many characters and high quality print; specialized memory units for information systems, possessing internal logic and the possibility of storage of large volumes of information, including photoscopic, associative, with internal logic; retrieval devices on continuous carriers (microfilms) and discrete carriers (microphoto cards, magnetic cards, etc.); technical means for mechanization and automation of all stages of information processes.

A considerable place in this volume is given to questions on organization of industry and improvement of technological processes on output of urgent information publications with wide application of methods of the classical printing industry and reproduction technology, the use of latest models of typesetting typewriters with stored control for manufacture of original mock-ups suitable for direct reproduction, application of xerographic, electronic equipment for urgent manufacture of printing forms, and also questions of the use of high-speed cylinder and offset machines for printing information.

Material is given on the use of highly productive brochure equipment.

THE ROLE AND PLACE OF THE MACHINE IN SCIENTIFIC AND TECHNICAL INFORMATION

A. I. Mikhaylov

The "...all possible assistance to further strengthening of the role of science in the building of Communist society... standard statement of scientific and technical information, and the whole system of study and propagation of domestic and foreign advanced experience" provided for by the program of the CPSU found its reflection during the last few years in a number of resolutions of the Central Committee of the Communist party of the Soviet Union and the Council of Ministers of the USSR directed towards creation in our country of a standard system of scientific and technical information.

At the 23rd Congress of the CPSU comrade A. N. Kosygin said: "Technical progress in the national economy and successes of science to a large extent depend on a well supplied system of information about results of scientific investigations conducted in this country and abroad, about the achievements and new methods of production, and about inventions and proposed innovations.

We must create in this country a highly effective state system of scientific information".

Recent years have been distinguished for active activity in the field of development of scientific and technical information in all links of the national economy.

Transition to a new system of leadership and planning of the national economy requires reconstruction also in the region of state leadership of scientific and technical information. The resolution of the Council of Ministers of the USSR from 29 November 1966 is a program of works directed towards creation of a state system of scientific and technical information most fully responding to further development of progress in our country.

During the last few years there has been scanned a wide front of development of scientific information. This finds confirmation in a huge number of publications and in a large number of international, regional, and national scientific conferences, symposia, conferences, etc. The main direction of these investigations is the search for ways to speed up processes of information activity.

In spite of successes attained in the region of scientific investigations and design developments, in our days the need has become evident critically to estimate results in the region of development of scientific bases and new technical means both from the point of view of satisfaction of appearing requirements of consumers of information and from the point of view of the overall solution of the whole information problem.

A deficiency of scientific activity in the past (during the experience of the VINITI)[All-Union Institute of Scientific and Technical Information] was individual special problems, which sometimes hurt the development of general theoretical problems.

At present more and more attention is being paid to questions of development of the theory of processes of scientific information. And the basis of information activity should be the theory of processes, not the empirical and intuitive method.

We do not as yet have an acceptable name for the scientific discipline which studies the structure and properties of scientific information; the regularity of information activity; and the theory, history, and method and organization of information. The term narrow and inconvenient. It seems to us that the most suitable term to use would be the word "informatika" [no translation found]. A similar proposal is argued in an article published in No. 12 "NTI" [Scientific and Technical Information] in 1966. From now on we will use this term.

We put in the idea "informatika" – in this new discipline – the following contents: the study of the laws, methods, and means of collection; analytic and synthetic processing; storage; retrieval; and propagation of scientific information.

In informatika much has already been done to give contemporary machine technology a fitting place in scientific and information activity. However, the role and place of the machine in informatika is sometimes understood to be too simplified, and in investigations and experiments with the use of ETsVM not all trends have been sufficiently developed.

The Role of ETsVM in Information Processes (The Time Factor Not Quantitative Growth)

Formation of scientific and technical information as an independent scientific trend appeared as a result not only of the exponential growth of scientific and technical literature, but also, in addition to that (and perhaps, even mainly), as a result of peculiarities of development of scientific and technical progress. Contemporary science and technology have two characteristic peculiarities.

1. Ever increasing complexity of scientific and technical problems the solution of which is possible only through the efforts of a large pool of scientists and engineers of various specialties. These pools need to be provided with information. Providing it takes on more and more the character of queueing. Effective solution of this problem requires the application of means of mechanization and automation.

2. Fast reduction of periods of development, mastering, and introduction of various discoveries and inventions.

The president of the Academy of Sciences of the USSR, M. V. Keldysh, in the article "Natural sciences and their value for development of Weltanschauung and technical progress" ("Kommunist", 1966, No. 17) writes: "Those who say that the century of science and technology has just now started are not entirely correct. After all in all times material and technical progress was largely based on the development of science. The roots of the industrial revolution also lay in science. The last century was characterized by a very large number of the greatest scientific achievements. It is incorrect to think that earlier there were few discoveries in the field of natural science entailing great consequences in the region of material production and that we have only now entered the century of continuous discoveries.

However, there is one feature very characteristic of contemporary development of science and technology; the speed of practical use of scientific discoveries. It is possible to give a number of examples. The period from the discovery of the electric current (Galvani) to the creation of the first electric power station spans about a century. It took almost one hundred years to master this remarkable discovery, having hugh prospects. It is possible also to note that seventy years passed from the clarification of the role of mineral fertilizers in the feeding of plants (the middle of the last century) to their intense use. And they came into wide use only after the second world war. The discovery of nuclear fission of uranium was another story. Only three years passed from the moment of this discovery to the creation of a nuclear reactor, and to creation of the first atomic electric power station 15 years.

The very fact that our time is characterized by extraordinarily fast use of achievements of science makes all the more important good organization of scientific investigations and use of their results in production. Today not that country which first makes a new scientific discovery but the one which is better able to organize its fastest use in practice is ahead in realization of the industrial process".

This peculiarity can be shown in a number of examples. From the moment of discoveries which in the final analysis led to the appearance of photography (the first half of the XVIIIth century) to the introduction of the means of photography in practice 112 years went by. Development of means of telephone communication took 56 years, radio 35 years, radar 15 years, television 12 years, and the transistor but 5 years.

Such acceleration of the rate of investigations and developments requires a corresponding increase in the speed of information systems. It is clear that this problem can be solved only by way of wide application in information practice of means of mechanization and automation.

Consequently, contemporary organization of information activity must not only help the researcher to look into the rapidly growing "Himalayas of libraries" but also satisfy the requirements of users, emanating from many aspects of the development of scientific thought.

The history of the development of informatika gives us many examples of the futility of trying to solve the problem of scientific information by way of creation of complicated machines simulating information processes. A look back at the past in American practice is enough to convince anyone of that. Remember the unrealizable hopes connected with the creation of such information-retrieval devices as "Rapid selector", "File search", "Minicard" and others. In 1958 we attended an American exhibition organized by the MFD [International Federation for Documentation] conference in the United States, where there was represented a wide range of modern machines which were not widely used in information practice. But at the same time work on such devices was useful at least in two respects. It permitted understanding that the basic difficulties of mechanization and automation of information processes consist not in the absence of necessary technical means but in the fact the internal mechanisms of fulfillment of such processes have not been studied by man. And when we are not clear on the mechanism of fulfillment of an operation by man, we can only imitate this process but not simulate it.

Another useful result of works connected with creation of information-retrieval devices was accumulation of valuable scientific and technical experience used more and more in various spheres of the national economy.

Speaking of the place of the machine in information activity, it is impossible not to remember tne definite evolution of views of specialists with respect to machine translation.

If in the first half of the fifties many researchers flippantly promised literally in a few years to replace the human translator with a machine, then subsequently their groundless optimism was replaced by a sober understanding of the exceptional difficulty of the problem before us. Attempts immediately "to take the bull by the horns" and already now realize machine translation in practice were replaced by deep theoretical research. Now there will hardly be found a serious researcher who will begin to affirm that the problem of machine translation (in its strict sense) can be solved in the next 2-3 years. It is true that experts of the "brain" corporation, "Rand," who are specially engaged in the composition of forecasts of development of science and technology in the next 50 years express a rather optimistic point of view — they expect macnine translation to become a reality already by 1970. Regarding, however, an automatic information center, they predict its appearance only in the period between 1970 and 1990.

Consequently, there appears the question of our relationship to the problems of machine translation. We have before us the very complicated problem of machine automatic reviewing and indexing.

It is possible to conclude this from reports which were read at the UNESCO seminar on the problem of automatic reviewing and indexing held in Moscow in September, 1966.

The scientific aspect of the problem is realization of automatic reviewing and indexing in the strict sense of these terms. This means that the unprepared text of the scientific document — book, article, patent, etc., is reviewed. The procedures of automatic reviewing can be pictured, for example, in the following way:

a) translation of the text of the document into a certain formalized language;

b) exposure of the main subject of this document and expression of the given subject in the formalized language;

c) translation of the abstract from formalized to natural language.

If the problem of automatic reviewing is solved, (the most important and difficult problem here is algorithmic exposure of the main subject of the scientific document), then automatic indexing becomes practically realizable. For this it is sufficient to translate the abstract from the formalized language into another formalized language utilized in the information-retrieval system.

Solution of this bunch of problems is possible only after carrying out deep fundamental investigations in many branches of science — in linguistics, psychology, mathematical logic, semantics, semiotics, etc. These investigations should open slightly the curtain above the secret of human thinking, which presents general scientific interest. Such investigations are so complicated that there is hardly any reason to expect considerable results in the next few years. But there can be no doubt that the investigations must be expanded and deepened. It is necessary to pay attention to the fact that without solution of the problem of machine translation we will not solve the problem of automatic reviewing and indexing. For execution of this vital problem there are useful and important all methods, including automatic quasi-reviewing in all its varieties and automatic indexing of abstracts of scientific documents. Of course here in the first place it is necessary to consider such factors as comparative cost of processing of texts, obtained gain in time, etc.

It is necessary specially to stress that development and application of methods intended for solution of such purely pragmatic problems by no means signifies solution of the problem of automatic reviewing and indexing in its strict sense, although it will promote this.

Thus, the examined problem has two aspects — scientific and practical. Both these aspects are important, but successes in the solution of the practical aspect of this problem must not be taken for the solution of its scientific aspect. A clear understanding of this distinction is a necessary condition of the success of our further work in this important direction.

Solution of the problem of automatic readout of texts is expecially important in connection with this. Without automatic reading devices not only can no practical system of automatic reviewing and indexing be realized, but investigations in sufficiently wide scales in this region are impossible. It is necessary to remember that all proposals to use punched tape obtained from flexowriters, monotypes, etc., for machine input and other analogous proposals have no bearing on the problem of automatic reading of texts, althougn they are very useful in the practical plan.

The side of the problem of automatic reviewing and indexing is the problem of satisfaction of the information needs of scientists and engineers today, and not sometime in the future.

One of the trends of scientific investigations which must be developed before the problems of informatika can be solved is undoubtedly semiotics. reation of artificial formalized languages of science and technology and also languages of generalized programming is one of the most important conditions of automation of information activity. Solution of the central problems of structural linguistics, machine translation and decoding of ancient written language will make it considerably easier for us to solve the problem of automatic analysis of the contents of scientific documents.

Semiotics are useful and necessary. Those of you who are acquainted with reports of it can evaluate its achievements yourselves. In connection with this I would like to make just one remark. It is assumed that results of semiotic investigations will find practical application in information activity in the future, when the basic problems of automatic analysis of text have been solved.

Meanwhile, we cannot wait this time. The recent resolution of the government obliges us to create soon an effective state system of scientific and technical information. This is why it is absolutely necessary that semiotics already now take part in a number of concrete works on creation of information-retrieval systems and on the study of information needs of specialists, improvement of methods of processing documents, and regulating of information flows. This will not only unite our efforts in solving pressing problems of informatika but will also make the investigations of semiotics itself move purposeful.

One more important trend in informatika appeared in connection with wide application of machine technology. I have in mind study of a system of scientific publications. Until lately the opinion was spreading that the information crisis arose in connection with the stormy growth of scientific publications. Authoritative scientists repeatedly expressed the opinion that publication of scientific literature needed to be limited and regulated. As examples it is possible to refer to the well-known project of J. Bernal proposing replacement of the contemporary system of scientific journals or the recent appearance in "Izvestia" of academician V. A. Kargin calling for limitation of the number of published sources.

However, investigations conducted at present abroad obviously show that we barely know the internal regularities of a system of scientific publications. Being basic means of transmission of scientific information in time and over distance, scientific documents obey certain objective laws, the neglect of which makes it impossible correctly to plan information activity and all the more so to improve the system of publications.

Long ago attempts were made to study this question on the basis of calculation of quoted literature and tracing of connections between documents forming by means of bibliographic references. However, only after this method was reinforced by the use of ETsVM, did it begin to give perceptible results. We must scan the corresponding investigations both with respect to domestic and with respect to foreign publications. Study of the laws of scattering and aging of publications and creation of information systems based on the method of bibliographic combination of documents (in particular, indicators of quoted literature), are among the first steps which must be taken.

* *
*

The purpose of this report is to show the place of the machine and its role in information activity. It seems to us that none of us needs to be convinced of the fact that the road to speeding up all information processes can be built only by the machine-automation in combination with the intellectual labor of man, necessary in all cases. Formation of the requirements of such machine technology must be approached from scientifically well-founded positions. We are faced with two large independent, internally interconnected problems.

The first concerns centralized processing of scientific and technical literature.

Conditionally it is possible to concieve of this problem as an information system at the input of which is a flow of world literature. After processing at the output of this conditional system we have bibliographic-signal information and series of abstract journals with a system of various indicators.

Here the role of ETsVM and its value is distinctly seen.

An acute problem remaining is creation of automatic input of printed text and satisfactory output from ETsVM of ready printed form for reproduction without typesetting.

An exceptional place is occupied by magnifying, highly productive equipment for printing with and without typesetting. But clarity in these individual questions does not remove the need to develop the wide front of research works spoken of in the beginning.

The second problem concerns the entire group of complicated questions connected with the creation of information-retrieval systems. Here great is the role of ETsVM as a means of helping to conduct experimental retrieval works in large volumes and in shorter periods.

A number of experimental retrieval systems have already been created in our country in the last few years. All of them are built on different principles, i.e., they are fragmental.

Is it possible already today to consider that these works have completely cleared up the whole problem? It seems to us, not yet. But at the same time these works are accumulating very interesting facts which are difficult to reevaluate. The sum of these facts will surely help correctly approach the creation of a state, centralized, or coordinated information-retrieval system.

In conclusion I would like to stress again and again the extreme need to develop a wide front of theoretical and experimental works illuminating the problem of creation of information-retrieval systems.

It is possible to say with confidence that an increasing staff of specialists in the field of scientific information will make their contribution to the great 50th aniversary of the October Revolution.

BASIC TRENDS IN THE DEVELOPMENT OF INFORMATION-RETRIEVAL SYSTEMS

D. I. Voskoboynik, G. E. Vleduts and
V. S. Chernyavskiy

Development of information-retrieval systems (IPS) at present has taken a rather wide swing in connection with the growing growth of information with respect to all branches of sciences and technology and increasing requirements of consumers of information on periods of information service and also in connection with acceleration of the rate of technical progress. Several years ago consumers of information could be satisfied by retrieval of materials with the help of the simplest IPS [information-retrieval systems] up to nonmechanized IPS of the type of library-bibliographic classifications.

Now such systems can no longer satisfy the demands of scientific and engineering-technical workers. Volumes of reference and information funds considerably grew, and interbranch problems represented by multiaspect materials became ever more important. In connection with this there were complicated requirements for retrieval systems, which now have to select materials from huge arrays with much more exact regard for various semantic aspects of these materials. EVTsM technology was called in to help. Its high speed was generally known and had already given a number of practically perceptible results in various regions of human activity.

However, the first experiments with the use of EVTsM for the purpose of information retrieval showed that high speed alone is not enough. In order to carry out multiaspect retrieval with the greatest thoroughness and minimum noise it turned out to be necessary to develop special information-retrieval languages (IPYa) which would allow recording information in EVTsM in a form useful for algorithmic processing.

Already in the early works of V. P. Cherenin [1] and V. A. Uspenskiy [2] dedicated to theoretical analysis of ways of developing IPS, the authors assumed that formalized IPYa determine potential semantic and logic possibilities of IPS independently of various methods of their technical realization.

However, this fundamental position very important for development of all IPS problems became widely acknowledged only during the last 2-3 years, especially after the II All-Union Conference on Information-Retrieval Systems and Automatic Data Processing. In general, one should note that the period which passed after the II All-Union Conference is characterized not only by considerable growth in the number of works in the IPS region, but also by essential increase in their scientific level. In this period trends also developed which were totally unrepresented earlier. Therefore, we can now present a more or less whole and systematized picture of the whole front of domestic developments in the IPS region during the last few years.

In the development of information-retrieval languages from the very beginning it is possible to trace two branches: IPYa for factographic IPS and IPYa for documentographic IPS.

Let us consider in the beginning the first branch.

As was shown in the work of V. A. Uspenskiy, the way to create sufficiently rich formalized languages for recording facts of natural sciences is through development of the metatheory of these sciences. In it there are basic forms of objects and relations corresponding to the elementary ideas of the given discipline and methods of construction of complicated ideas from elementary ideas and the reasoning procedures applied in the given region are investigated. On the basis of created metatheories formalized languages are constructed as methods of recording facts with the use of the symbolism of mathematical logic. Formalized languages thus built can be used in systems automating not only information retrieval, but also processes of forecasting new facts in so-called "information-logic" systems.

Examples of particular formalized languages created on the basis of the above-mentioned metatheoretical-logical approach are formalized languages developed for geometry and organic chemistry. Information-logic language built on the basis of narrow predicate calculus by A. V. Kuznetsov, Ye. V. Paducheva, and N. M. Yermolayeva [3] for geometry was subsequently a successful object for investigations carried out by Ye. V. Paduchevoy [4] of the problems of translation from information-logic languages to Russian. G. E. Vleduts and V. K. Finn [5] developed information-logic language for structural organic chemistry. The mentioned formalized language was assumed as a basis of development for the field of chemistry of the big information-logic system. In this system units of information are properties of chemical compounds and processes of their mutual transformation, i.e., chemical reactions. For machine recording of the structure of organic compounds and equations of organic reactions there are used atomic (topological) linear recordings of chemical graphs, and also several different systems of so-called "filter recordings," reflecting with various degrees of detail the basic peculiarities of structure of compounds and chemism of reactions.

For the putting into the system of nonlinear chemical structural information there have been developed special systems of primary coding close to the nomenclature language and designations accepted in chemistry.

Leaning on the peculiarities of formalized language utilized in the system it was possible to formulate algorithms of solution of a number of information-logic problems from the field of chemistry, in particular, a probable forecast of reactivity and ways of synthesis of organic compounds.

Other examples of factographic IPS are the system for inorganic chemistry developed by A. L. Seyfer and his colleagues [6], in which the chief attention is focused on recording and retrieval of properties of compounds expressed in numerical form and also IPS for physico-chemical systems with two or more components based on formalized language of physical-chemistry phase diagrams.

During construction of information languages, forming the basis of factographic IPS, there was more or less carried out the above-mentioned metatheoretical approach to formalization of the language of the corresponding fields of science.

Another approach to construction for sufficiently broad fields of science and technology of information languages of the so-called "descriptor" type is based on study of vocabulary specific for the natural language of the corresponding branch. It is accepted practice to call terminological dictionaries thus compiled of words and word combinations grouped in classes of term-descriptors sufficiently close or equal in meaning "thesauri" or descriptor dictionaries.

In thesauri, besides exposure of the relationships of synonymy and homonymy of terms, there are also exposed semantic relationships existing between descriptors. These relationships reflecting objective relationships taking place in the examined subject field are customarily called "basic" relationships according to the terminology introduced by V. S. Chernyavskiy.

Descriptor languages are widely used in documentographic IPS. An example of a descriptor dictionary for large-scale documentographic IPS in the field of applied chemistry and the chemical industry is the thesaurus developed under the leadership of V. B. Margaritov.

The retrieval forms of documents on descriptor information-retrieval languages without grammar have the form of simple sets of descriptors. Use of such languages essentially simplifies algorithmization of the process of translation from natural to formalized language, which in this case essentially boils down to word for word comparison of the text of the document in natural language with the thesaurus; difficulties connected with recognition of homonymy are easy to solve by lexical analysis of contextual encirclement of homonymic words. This was shown by V. S. Chernyavskiy and his colleagues on an example of translation of texts of abstracts on electrical engineering from Russian to descriptor language without grammar. A language of such type, with the establishing of basic relationships between descriptors, is assumed by V. S. Chernyavskiy, D. G. Lakhuti and E. S. Bernstein [7] as a basis of the "Pusto-Nepusto" experimental IPS developed by them for the field of electrical engineering. In particular, the "Pusto-Nepusto" IPS in which documents are automatically indexed according to the above-mentioned principle, is realized with the help of the "Minsk-22" EVTsM.

There have been developed information languages of the descriptor type for IPS intended for separate, more or less wide regions of technology, in particular, for a number of subfields of radio electronics, tractor construction, machine-tool building, and others. In the Central Institute of Patent Information thematic dictionaries for corresponding divisions of the patent fund are compiled simultaneously with development of the overall experimental system of machine translation and automatic indexing.

As we have already mentioned above, retrieval samples of documents in descriptor languages without grammar have the form of simple sets of descriptors. Because of the absence in these languages of grammatical means such retrieval samples cannot reflect contextual relationships into which semantic units corresponding to descriptors in concrete text enter. This deficiency leads during algorithmic retrieval to increase in the percentage of unnecessary delivery, i.e., to rise in the number of issued IPS documents not satisfying the inquiry. To decrease this "retrieval noise" descriptor languages are given grammar elements, the simplest of which are so-called "role indicators," which are supplied to individual descriptors in retrieval form for the purpose of defining more accurately their menaing in the examined context. A. V. Sokolov proposed using three role indicators the first of which would separate one or more main descriptors corresponding to the main subject or subjects of consideration. This role indicator distinguishes the main descriptors playing the role of "subjects" from the remaining descriptors fulfilling explicative functions of "adjectives." Furthermore, role indicators of negation (absence) and multiplicity are used.

At the Institute of Cybernetics of the Academy of Sciences of the Ukrainian SSR under the leadership of E. F. Skorokhod'ko [8] an IPS has been developed for the field of computer technology based on descriptor language, in which well-developed grammatical means are used. In this language from a comparatively small number of base (elementary) terms with the help of a limited number of base related themes fulfilling the role of one-place and two-place predicates, more complicated terms of different rank are constructed. "Scannings" of complicated terms thus obtained reflect the structure of definitions of complicated ideas with the help of initial elementary ideas corresponding to the base terms. Scannings of terms assign basic semantic relationships between terms, on the basis of which semantic correspondence between retrieval patterns of documents and inquiries are established. Related terms are used also to transmit textual relationships between terms in retrieval patterns.

For the field of synthetic organic chemistry N. A. Stokolova and D. G. Lakhuti [9] built a descriptor language with grammar in which syntactical connections are expressed by the method of "standard phrases." A standard phrase is a means of type of multiplace predicate the places of which are filled with term-descriptors. Every form of standard phrase is used for recording a certain type of information and the place, occupied by the term in the standard phrase and strictly determines its function in context. To record in information language texts constituting titles of abstracts of articles from the examined field it turned out to be sufficient to use three forms of standard phrases of various degrees of complexity. The basic relationships between terms are expressed in the form of classification diagrams; several different forms of relationships are used. Experiments in thematic retrieval of information with the use of this language showed that very high values of coefficients of accuracy and thoroughness of retrieval are attained in this way.

It is necessary to stress that from the point of view of operational qualities of IPS, besides the semantic force of information languages utilized in them, peculiarities of methods of realization of these systems and, in particular, such methods of organization of information arrays as the direct and inverse methods are important. From this point of view one should note development of the method of associative programming, which permits realizing the associative-address method of organization, combining a number of positive features of both the direct method and the inverse method. This method is now realized in the "Setka-3" IPS, which services a number of branches of radio electronics, and in IPS on materials of cardiovascular surgery.

To evaluate the influence of various structural elements of IPS on their operational parameters, experimental studies comparing the efficiencies of various systems have positive value. Such a work, in which an experimental method was worked out and the efficiency of traditional IPS (based on library methods of classification and cataloging) was compared with that of descriptor IPS was carried out at the Leningrad Institute of Culture on the initiative of A. V. Sokolov [10].

To further improve IPS it is necessary to intensify theoretical research, which will allow comprehending, correctly estimating, and improving the means of expression of information languages. This, of course, is impossible without thorough comparison of the peculiarities of the means of expression of artificial and natural languages. From this point of view works already mentioned on automatic indexing have great value, i.e., automatic translation from natural languages into various kinds of information languages, and also works close to them on automatic reviewing and automatic composition of subject indicators. An important general trait uniting all these processes is compression of information, i.e., transition to the most important characteristics of the meaning of the document.

The most interesting results in the field of automatic indexing (cataloging) are obtained during overall use of various principles, in particular, during combination of means of the following three different types: 1) lexical means in the form of precomposed lists of words (terms) significant or insignificant for the given field, 2) statistical data on the particularity of different words (terms) in the examined text or in the totality of texts from the given region and 3) results of syntactical or another kind of formal structural-linguistic analysis utilized for identification of semantic categories.

In the first works on questions of automatic reviewing [11, 12] there were investigated diverse variants of purely statistical procedures of sampling weighted phrases of reviewed text. Recently there have been developed methods combining statistical methods with the other means mentioned above. Thus, in the work of V. M. Gorobtsov [13] on automatic cataloging along with frequency considerations there is considered entry of words both into descriptor dictionaries of the whole field and into descriptor dictionaries compiled for texts pertaining to various subject classes and also certain grammatical characteristics of words. In the end, the text is classed under a certain subject heading with a known degree of probability.

In the work of N. I. Styazhkin and his colleagues [14] after translation of the text of the document into descriptor language the phrases containing the greatest number of descriptors encountered in the title of the article and encountered jointly with the descriptors of the titles are selected: in this way there are obtained "author's abstracts"[1] of acceptable quality in the sense that deviation between them and "abstracts" composed by people (by way of sampling phrases taken from reviewed text) is not too great.

[1] This term is used here in the sense of a "paper" composed automatically.

In the work of B. V. Yakushin [15] the basis for algorithmization of composition of subject recordings is a certain list of base terms fixed for the given field. For exposure in the text of other words explaining these base terms and forming together with them "calling pairs" which are used as subject recordings, grammatical (syntactical) criteria are attracted. By this method subject recordings fully comparable with subject recordings composed by indexers are obtained.

Along the line of use of linguistic means is the work of I. P. Sevbo [16], in which the results of a complete syntactical analysis in a certain sequence of phrases are used: here not only separating of certain semantically weighted sections of phrases (nominal groups) occurs, but procedures of their unification in chains connected in meaning also appear; the result (an abstract of the annotation type) is a list of names on which the text is discussed.

Comparing the level of domestic and foreign works over the whole complex of questions pertaining to formalized languages of science and technology, and, in particular, the state of theoretical research in this field, it is possible to note that we have serious achievements creating favorable prerequisites for successful advance forward.

Further progress in development of both methods of construction of rich enough information languages and methods of translation into them from natural languages and back requires essential deepening, expansion, and approach of fundamental structural-linguistic and logic investigations. From this point of view research in creation of semantic information theory is paramount; important steps in this direction are made in works of Yu. A. Shreyder [17] by way of generalization and deepening of the thesauris concept.

It is necessary to allot much attention to investigations directed towards automation of individual stages of processes of creation of IPS, in the first place automatic composition of thesauri. For these purposes it is necessary to more widely apply statistical methods to data stored in big information funds.

Extraordinarily important problems appear in connection with providing IPS with machine technology. The main difficulty here is caused by the comparatively low class of series domestic EVTsM: insufficient volumes of storage units, poorly developed parallelism of devices, and low reliability – especially of external devices.

These deficiencies lead to the need rigidly to save machine operations and memory volumes in the process of programming, which in turn prevents standardization and automation of programming. In the end programming of even not very complicated algorithms is turned into time-consuming work. At present programming is frequently the cause of prolonged delays in carrying out necessary experiments and in realization of already developed IPS.

It will be possible to change this position, if industry schedules EVTsM, at the level of contemporary average world standards.

Although we cannot at present completely answer the question of optimum parameters of EVTsM intended for information or semantic purposes, accumulated experience permits formulating certain unconditionally necessary requirements.

Speed is not a critical parameter, but it is desirable to bring the number of operations per second up to at least 100 thousand.

Capacity of storage units – fast store should be brought up to 32 thousand words, and the capacity of external memory should be brought up to several billion bits.

Types of storage units – access to external memory should be facilitated and it is desirable to have drums or disks or sufficiently convenient tape units. External memory in which at least with respect to one process – readout or access time would be the same order as in fast store is very useful. All these devices are developed at domestic enterprises, but their introduction and issue lag.

Input and output – the presence of many simultaneously operating input and output devices is absolutely necessary, in particular readers and multialphabetic high-speed output units of the photosetting-machine type.

Parallelism – independence and parallelism of functioning of all machine units are necessary and also the possibility of operating in the multiprogramming mode.

These will be definitized or modified with the accumulation of practical experience. However, one should recognize that the process of accumulation of this experience is extremely slow. We have a certain number of successfully operating experimental systems, but as yet not regularly functioning big information-retrieval services. And meanwhile, only in the process of industrial exploitation is it possible to organize comparative investigations and to work out optimum criteria.

Let us try, however, to formulate certain general positions, proceeding from existing domestic and foreign experience.

For effective exploitation of IPS they must be used for simultaneous solution of two problems: 1) for selective (address) announcement of consumers on new entries on the assigned subject; 2) for retrospective retrieval with respect to inquiries. It is necessary to design systems which could provide the serviced circle of consumers with all forms of information necessary to them. For this purpose automated IPS should, besides solving the above-indicated problems also be widely used for composition of various types of signal bulletins or subject indicators (in particular, of the permutation type). Automated IPS, realized on EVTsM, must also produce punched card variants of small retrieval systems useful for reproduction and use at places with application of simple means of mechanization.

In the process of realization of large-scale branch systems and information services an important role must be played by the ministries and branch information centers subordinate to them. Disposing big material resources, ministries can make a decisive step for transition from experimental and test systems to large-scale operational services. All branch services must be intelligently combined with the centralized system of processing scientific and technical literature in VINITI and scoop from VINITI materials for filling branch IPS in a form convenient for this. It is clear that VINITI, as head institute, is ready actively to participate in the process of designing branch services and must generalize and spread experience accumulated during designing and exploitation.

We must always remember the indication of Comrade A. N. Kosygin, Chairman of the Council of Ministers of the USSR, made by him in a speech at the XXIII Congress of the CPSU, about the need to create a highly effective state harmonious and reliable system of scientific information. An interconnected system of branch services, servicing in the beginning narrow and then wider and wider fields of science and technology, gradually has to satisfy all interests of scientific and engineering-technical workers. Complexes of machines supplied with a branched network of lead-in (reading) and lead-out devices placed directly in places of generation and consumption of information, in scientific-research and research-design establishments, and united by means of communication with central devices will fulfill more and more diverse forms of data processing. Storing in their memory units a whole mass of introduced information, they will deliver it to consumers in accordance with thematic requisitions formulated by them, which will be definitized as a result of constant feedbacks. Information should be delivered in the form of compressed summaries, but on demand of the consumer and in the form of detailed abstracts or detailed factographic references. Furthermore, periodically (in accordance with thematic profile) or on demands, scientists, engineers and leaders have to obtain thematic surveys or specialized indicators, and products of logical and statistical processing of accumulated information more complicated in perspective up to forecasts of facts or hypotheses.

Results obtained up to now, familiarization with many of which awaits us at sessions of sections, open fully real prospects of achievement of outlined targets.

Bibliography

1. Cheronin V. P. Nekotoryye problemy dokumentatsii i mekhanizatsii informatsionnykh poiskov (Certain problems of documentation and mechanication of information searches). M., In-t nauchn. inform., 1955.

2. Uspenskiy V. A. K probleme postroyeniya mashinnogo yazyka dlya informatsionnoy mashiny (The problem of construction of machine language for a data-processing computer). In the collection "Problemy kibernetiki". Vyp. 2, M., Fizmatgiz, 1959, str. 45.

3. Kuznetsov A. V., Paducheva Ye. V., Yermolayeva N. M. Ob informatsionnom yazyke dlya geometrii i algoritme perevoda s russkogo yazyka na informatsionnyy (An information language for geometry and an algorithm of translation from Russian to information language). In the collection "Lingvisticheskiye issledovaniya po mashinnomu perevodu". Soobshcheniye otdela mekhanizatsii i avtomatizatsii VINITI. Vyp. 2, M., 1961, str. 40.

4. Paducheva Ye. V. Nekotoryye voprosy perevoda s informatsionno-logicheskogo yazyka na russkiy (Certain questions of translation from information-logical language into Russian). "NTI", 1964, No. 2, 20.

5. Vleduts G. E., Finn V. K. Problematika sozdaniya mashinnogo yazyka dlya organicheskoy khimii (Problems of creation of machine language for organic chemistry). In the book "Soobshcheniye laboratorii elektromodelirovaniya". M., VINITI, 1960, str. 67.

6. Seyfer A. L., Shchurova S. S., Polyusuk Yu. A. Avtomaticheskaya informatsionno-poiskovaya sistema dlya neorganicheskikh soyedineniy (An automatic information-retrieval system for inorganic compounds), "NTI", 1963, No. 10, 26.

7. Bernshteyn E. S., Lakhuti D. G., Chernyavskiy V. S. Nekotoryye voprosy postroyeniya informatsionno-poiskovykh sistem (Certain problems of construction of information-retrieval systems). "NTI", 1963, No. 1, 31.

8. Gryazkukhina T. A., Pshenichnaya L. E., Skorokhod'ko E. F. Sistema informatsionnogo poiska (An information-retrieval system). Kiyev, "Naukova Dumka", 1964.

9. Lakhuti D. G., Stokolova N. A. O zadache poiska khimicheskikh referatov po zaglaviyam (The problem of retrieving chemical abstracts by titles). "Doklady na konferentsii po obrabotke informatsii, mashinnomu perevodu i avtomaticheskomu chteniyu teksta". Vyp. 1, M., 1961.

10. Sokolov A. V. Issledovaniye poter' informatsii i informatsionnogo shuma v deskriptornykh informatsionno-poiskovykh sistemakh (Investigation of information losses and information noise in descriptor information-retrieval systems). "NTI", 1965, No. 12, 23.

11. Purto V. A. Ob avtomaticheskom referirovanii na osnove statisticheskogo analiza teksta (Automatic reviewing on the basis of statistical analysis of text). "Doklady na konferentsii po obrabotke informatsii, mashinnomu perevodu i avtomaticheskomu chteniyu teksta". Vyp. 1, M., 1961.

12. Agrayev V. A., Borodin V. V., Glebskiy Yu. V. O nekotorykh metodakh avtomaticheskogo referirovaniya (Certain methods of automatic reviewing). "Uchenyye zapiski Gor'kovskogo gosudarstvennogo universiteta im. N. I. Lobachevskogo". Vyp. 66, seriya – filologiya, prikladnaya lingvistika i metodika. Gor'kiy, 1963, str. 73.

13. Gorobtsov V. M. Statisticheskiy metod indeksirovaniya i poiska informatsii (A statistical method of indexing and retrieving information). Trudy simpoziuma SEV "Kompleksnaya mekhanizatsiya i avtomatizatsiya protsessov obrabotki, poiska, vydachi i peredachi na rasstoyaniye nauchno-tekhnicheskoy informatsii". M., 1965.

14. Monastyrskiy I. M., Pevzner B. R., Styazhkin N. I. Metod deskriptornogo avtomaticheskogo referirovaniya literatury po stankostroyeniyu i printsipial'nyy algoritm ego realizatsii na EVTsM (A method of descriptor automatic reviewing of literature on machine-tool building and the fundamental algorithm of its realization on EVTsM). "NTI", 1964, No. 2, 28.

15. Yakushin B. V. Algoritmicheskiy metod vydeleniya predmetnykh ponyatiy dlya sostavleniya ukazateley (metod nazyvayushchikh par) (The algorithmic method of deriving subject ideas for composition of indicators (method of calling pairs)). "NTI", 1963, No. 7, 12.

16. Sevbo I. P. Ob odnom metode avtomaticheskogo annotirovaniya teksta (One method of automatic annotating of text). "NTI", 1965, No. 8, 32.

17. Shreyder Yu. A. Ob odnoy modeli semanticheskoy teorii informatsii (One model of the semantic theory of information). In the collection "Problemy kibernetiki". Vyp. 13, M., "Nauka", 1965, str. 233.

PREPARATION OF ENGINEERING AND SCIENTIFIC CADRES WITH RESPECT TO MECHANIZATION AND AUTOMATION OF INFORMATION WORKS

Doctor of Technical Sciences Prof. I. I. Petrov and
I. B. Pochkay

The most important condition necessary for successful development of scientific and technical information in the country is automation of various information processes based on computer technology and other contemporary means of automation. In connection with this questions of educating specialists in mechanization and automatic data processing have become very important. Questions of training cadres were given especially great attention at the XXIII Congress of the CPSU. In the current report of the Central Committee of the CPSU it is stressed that these questions have to be advanced to the level of general political problems of the party and state.

Up to 1965 there was in general no preparation of specialists of such a profile in the USSR, and by this time about 100,000 people have worked in the system of scientific and technical information. It is natural that unproductive "manual" methods have predominated in information services and that processes of information service were mechanized and automated very slowly. Taking all these circumstances into consideration, the Ministry of higher and special secondary education of the USSR decided in 1965 to create in the higher school system specialty No. 0640 – "Automation and mechanization of processes of processing and delivery of information." This specialty is offered in four higher educational institutions of the country: the Kuybyshev Polytechnical Institute, the Tomsk Institute of Radio Electronics, the Sevastopol' Instrument-making Institute, and the Tallin Polytechnical Institute.

In the first course of all these institutes 100 people are taught, and in the second course 75 people. Certainly, this is small. In order to expand preparation of engineers for the information organs of the country, it is necessary to ask the Ministry of higher and special secondary education of USSR to offer in the near future specialty No. 0640 in at least 6 higher educational institutions of country, including the higher educational institutions of Moscow, Leningrad, Kiev, Sverdlovsk, and in other cities, where there is specially sensed a sharp need for specialists in automatic data processing and where there are scientific and pedagogical cadres in the field of automatics, computer technology, and technical cybernetics. The solution of this problem is of interest not only to the information organs of the country but also to the Ministry of Instrument-making, Means of Automation, and control systems of the USSR, which is assigned development and production of special technical means and systems for processing, storage, and retrieval of scientific and technical information; copying-duplicating equipment; means of microfilming; typesetting-typewriters; computers; and library equipment. Here it should be especially stressed that the specificity of specialty No. 0640, the basic disciplines of which are based on computer technology, automatics, electronics, communication engineering, etc., requires for organization of laboratories for this specialty scarce and complicated equipment. Therefore, with the offering of this specialty in higher educational institutions it is necessary to provide their corresponding material base, allowing for this sufficient means and funds, including the purchase of imported equipment.

A very important problem also requiring immediate solution is the improvement of the curriculum of specialty No. 0640, and programs of discipline entering it. Many of the tenets of this plan are due to novelty, and the experience of its compilers turned out to be insufficiently founded.

At present the Scientific and Methodical Council on Automation of Industrial Processes of the Ministry of Higher and Special Secondary Education of the USSR recommended proceeding from the following general principles in developing new curricula on automation.

1. Disciplines of physicomathematical and engineering cycles must not contain obsolete information, methods of calculations and investigations historically composed, but having lost practical value, and material duplicating other disciplines. Programs of these disciplines must be based on the attained level of natural (mathematics, physics) and applied technical sciences.

2. A mathematics course must reflect the bases of mathematical logic, the principles of calculus of variation, probability and information theories, and other divisions needed to increase the mathematical level of the specialist in automation. These divisions must not be introduced into the course as simple additions to the existing complex of mathematical questions, but must be an organic part of the whole course of higher mathematics. It is especially important that the study of higher mathematics be conducted on the basis of application of computer technology. In general, provision should be made for using computer technology in all disciplines of both physicomathematical and engineering cycles and the special cycle.

3. The contents of disciplines in theoretical mechanics, strength of materials, theory of machines and mechanisms, descriptive geometry and drawing, and other general-engineering disciplines not directly related to specialties in automation should be radically examined. It is necessary somewhat to reduce the nomenclature of these disciplines, decrease the volume of certain of them, and thoroughly examine their contents.

4. It is necessary to strengthen in all possible ways the general and special education of future automation engineers in electrical-engineering disciplines, electronics, and the general theory of automatic control.

5. It is desirable in forming the nomenclature of special disciplines not to crush it in a whole series of small courses, but to provide in the curriculum for fundamental study of enlarged special disciplines embracing the basic questions of the specialty. Such an approach will exclude duplicating of materials in programs of separate disciplines and increase the quality of preparation of automation specialists of a wide profile. It is necessary to provide for a certain sequence in the study of special disciplines so that the study of the basic special disciplines precede the study of narrower special disciplines.

It is necessary to provide for a certain sequence in the study of special disciplines so that the study of the basic special disciplines precede the study of narrower special disciplines.

6. A very important problem in the examining of the contents of disciplines of automation curriculums is elimination of scholasticism in the study of these disciplines. This scholasticism is caused by the tendency to describe any phenomenon or process with only a mathematical formula without any explanation of the essence of the phenomenon or process. Departure from physics permits the student to start to perceive the physical process through the prism of mathematical expression, not penetrating into the essence of the phenomenon, and to become helpless if this process is modified.

Unfortunately, in the curriculum of specialty No. 0640 affirmed by the Ministry of Higher and Special Secondary Education of the USSR, the expounded principles were not adequately taken into consideration, which made the plan far from perfect.

Thus, the plan contains the discipline "Hydraulics and hydraulic machines" (123 hours), which is not strictly necessary for the profile of the specialist being educated and its exclusion from it is not detrimental. There is insufficiently founded removal from the curriculum of the courses "Strength of materials," "Theory of machines and mechanisms" "Machine parts" and leaving in it the course "Theoretical mechanics." Here, as is done in the curriculum of specialty No. 0606 – "Automatics and telemechanics" and in curriculums of other automation specialties, it would be more expedient to combine all these disciplines into a single course, "Mechanics," and mainly stress questions of dynamics, the theory of elasticity and the theory of oscillations in it.

No provision is made in the plan for the study of one of the most important basic disciplines for automation engineers, namely: "Theoretical bases of electrical engineering," containing expanded theory of electric circuits at the expense of a certain reduction in field theory. Instead of it into the plan there is introduced a course in "General electrical engineering," which in no case can be considered founded. This error must be corrected and a large number of training hours must be assigned to the study "Theoretical bases of electrical engineering."

Absent from the curriculum are disciplines important for future automation engineers, such as the "Theory of Automatic Control and Checking" and "Mathematical Bases of Cybernetics," which leaves serious gaps in the plan of specialty No. 0640.

Special disciplines are in especially bad shape. They are excessively crushed and to a considerable extent duplicate one another; therefore, the composition of programs in these disciplines is very difficult. Examples of such disciplines are "Technological processes, machines, and apparatuses of scientific and technical information" (247 hours), "Means of reproduction of scientific information" (140 hours), "specialized additional units and devices of data processing computers" (132 hours), "Systems based on punching technology and electronic computers" (85 hours), "Construction and exploitation of punch-out computers (50 hours) and several others.

In order to correct these deficiencies it is apparently necessary to enlarge special disciplines and provide for the study in them of such questions as information-retrieval systems, automation of technological processes, and others. The sequence of study of special disciplines is not maintained in the curriculum. An example is the course "Bases of scientific and technical information," the study of which starts only from the 8th semester, and highly specialized disciplines start to be studied from the 6th semester, that is, they precede the basic special discipline. This is an essential deficiency of the curriculum, and it must be corrected.

Thus, the curriculum of specialty No. 0640 now in effect is imperfect and needs essential corrections. Inasmuch as students learn specialty No. 0640 only in the second course, there is time for these corrections. It is necessary to ask the Ministry of Higher and Special Secondary Education of the USSR to examine the curriculum of specialty No. 0640 and correct it as necessary.

Let us now turn to the question of preparation of scientific cadres for the information services of the country.

At present scientific cadres are taught automatic data processing in the USSR only at VINITI. In 1959 it began to offer post graduate work in the three following specialties: "Scientific and technical information," "Computer technology" and "Computer mathematics." All these specialties are being studied by 62 graduate students, including 36 people studying "Scientific and Technical Information." In the past 6 years 14 people have completed post graduate work, 8 of them in 1966. This is very small. Apparently, in the near future it will be necessary to offer specialty No. 0640 for preparation of engineers (which was mentioned above) not only in big higher educational institutions of the country but simultaneously to organize preparation of graduate students in automatic data processing in these higher educational institutions.

A large potential reserve for preparation of science candidates in the field of information is the so-called "competitors," a number of leading specialists of information services. Thus, in 1965 out of the number of VINITI workers alone and especially from its scientific-research subdivisions, more than 25 engineers started to work on dissertations, using VINITI scientific-research laboratories as an experimental base.

The educating of scientific cadres in the field of automation of information processes is considerably deterred by the informing of specialists about the scientific set of problems of this new branch of knowledge. This set of problems is very interesting and many-sided. It is formed at the junction of many sciences and uses both achievements in the field of semiotics, and achievements in the field of automatics and telemechanics, computing and electronic technology, cybernetics, and other sciences. It was very timely to compose and issue a special scientific and methodical aid with an account in it of the basic problems of scientific information and problems of automation of information processes and widely to diffuse it among the specialists of information services of the country.

Considerable difficulties arise in the selection of scientific leaders of graduate students and competitors. It is necessary more widely to attract to such leadership not only scientists working in organs of information but also scientists from various kinds of scientific research institutes and higher educational institutions, the thematic directivity of which is close to the problems of automatic data processing (institutes of cybernetics, automatics, and telemechanics, computer technology, etc., and also the corresponding departments of higher educational institutions).

At present problems of increasing scientific and engineering qualification of cadres occupied in the field of development, mastery and exploitation of information-retrieval systems and mechanization and automation of processes of processing of scientific and technical information are becoming very important. It is necessary more widely to practice the organization of constantly operational and short-term courses, scientific and engineering cadres and also leading cadres of institutes of information, using the experience of the VINITI. At those higher educational institutions of country where specialty No. 0640 is offered, it is expedient to create courses to increase the qualification of specialists working in the field of automation and mechanization of information processes. It is also necessary to more widely attract scientist-candidates and doctors of sciences working or having education in the field of scientific information to the reading of lectures and teaching in higher educational institutions, in various courses, and at various seminars.

THE LOGIC OF DESCRIPTOR RETRIEVAL SYSTEMS

V. S. Chernyavskiy

During construction of retrieval systems, including so-called descriptor retrieval systems, it is always necessary to make a large number of different assumptions, which usually are not clearly formulated. These assumptions, which in the most essential manner determine both the structure and the properties of created systems, cannot be derived from any theories developed up to the present time and at the same time, as far as it is now possible to judge, cannot be confirmed or refuted by no matter what experiment different from direct experimenting with retrieval systems built on their basis. Therefore, the most natural evaluation of retrieval systems is one which in evident form operates with assumptions on which they are based and rests on more or less considerable experience of their exploitation. In this article such analysis is conducted for two systems of the "Pusto-Nepusto" class — the systems "Pusto-Nepusto"-4," developed by Bernshtein and "Pusto-Nepusto-2," developed by Lakhuti.

For sources of that group of retrieval systems to which there belong, in particular, retrieval systems of the "Pusto-Nepusto" class, there lies an idea of fundamental importance for the first time expressed and realized, as far as can be judged, by Mortimer Taub in his "Uniterm" system. This idea consists in the fact that in natural language it is possible to separate certain "significant" words that with a completeness, sufficient for the purposes of information retrieval, the contents of documents and inquiries will be transmitted by a disordered set of "significant" words entering them. In other words, the Taub idea consists in the fact that for retrieval purposes it is sufficient to consider that part of the contents of documents and inquiries, which is transmitted by their dictionary composition.

Both this idea itself and various modifications of it have provoked and are till now provoking numerous objections. These objections basically boil down to affirmation that retrieval systems not taking textual relationships between words into account cannot be effective. As an argument there are usually given various examples, such, let us say, as "influence of dyes on bacteria" and "influence of bacteria on dyes." These examples have to show that dictionary composition of sentences can be the same, and at the same time, if one of them is considered an inquiry and the other, let us say, as the title of a document, then the document should hardly be issued on demand.

In spite of the apparent convincingness of such objections, at present the successfully exploited retrieval systems of the "Uniterm" type, are well known, and this simple fact indicates that the matter is by no means examples contradicting the Taub idea or any other idea, but in whether such examples are encountered in concrete conditions of functioning of the retrieval system and in a quantity noticeably lowering its effectiveness. Thus, one may assume that in spite of the existence of contradicting examples and possible objections, the experiment confirmed the correctness of the Taub idea, it goes without saying only for those concrete conditions in which "Uniterm" type systems are exploited, so that there are no bases to ascribe to this idea great universality. What has been said well illustrates that very important circumstance that now in the discussed set of problems speculative reasonings can be only the initial point of investigation, but not its replacement; reliable conclusions can be drawn only on the basis of experiment.

In accordance with the simplest treatment of the Taub idea, if it is taken literally, the document would have to be issued in answer to those and only those inquiries, the dictionary composition of which coincides with its dictionary composition. In such a form the

Taub idea was not grasped by anyone and was not realized, at least, in systems which are appropriately called systems. From the very beginning this idea was transformed in that direction, which for delivery of a certain document is sufficient so that it included all terms from which the inquiry is built.

This additional assumption is not evident and cannot be obtained deductively as a result of generally significant evident positions. Moreover, it is easy to think of a situation in which it will be incorrect, since relevance or irrelevance is not an immanent property of the document-inquiry pair and can essentially depend on the conditions of functioning of the system. Such a situation can be caused, in particular, by the use as information objects of multitheme documents and documents large in volume. In these conditions it can happen that a small document wholly dedicated to the subject of the inquiry is relevant, while a large volume concerning the subject of the inquiry is not relevant and this in spite of the fact that in a large document to the subject of the inquiry there will be assigned as much place as in a small one. The possibility of such a situation is anticipated in particular by Harvard University researchers (the United States), when they offer for calculation of the relevance of a document a formula according to which relevance turns out to be inversely proportional to the volume of the retrieval pattern of the document.

This assumption can be opposed on the same grounds as the Taub idea, namely, it is possible to give as a contradicting example some artificial document or one encountered in practice, including the whole dictionary composition of an inquiry the relevance of which nonetheless is more than doubtful. An example of such an inquiry is "voltage of generators utilized on submarines" and the title of the document "repair of high-voltage generators utilized on submarines." And nevertheless, in spite of possible objections and the existence of examples contradicting the discussed assumption, it is confirmed by successful functioning of systems of the "Uniterm" type; it is confirmed, of course, only for those concrete conditions in which these systems function.

Systems of the "Uniterm" type are based on a third assumption — namely, on the assumption that entry of all the terms of the inquiry into the dictionary composition of the document is not only sufficient but also necessary for their relevance. Like the other two assumptions already examined by us, the third assumption is not evident and it is easy to come up with examples contradicting it. These examples break down into two groups essentially differing from each other.

The first group can be represented by an inquiry in which there is an adjective or another word limiting its subject. Let us assume that, for example, there is a text relevant to the inquiry "production of transformers." Then it can happen that it is also relevant for the inquiry "production of large transformers" in spite of the fact that the word "large" does not appear in its dictionary composition. Practice, however, shows that such examples, though they be completely real, do not noticeably lower the efficiency retrieval systems of the "Uniterm" type and therefore cannot be the cause of transition to systems of another type.

It is an entirely different story with examples of the second group since difficulties connected with them can no longer be disregarded. Let us consider as an illustration the inquiry "exploitation of high-voltage equipment" and a document under the heading of "repair of small-oil circuit-breakers." Not one term of the inquiry enters the title of the document and nevertheless, even without turning to the text of the document, it is possible to say with confidence that it is relevant to the inquiry since questions of repair pertain to exploitation, and small-oil circuit-breakers pertain to high-voltage equipment. Practice shows that such cases cannot be disregarded, since this would lead to unacceptable losses of information. Therefore, the Taub principles must be essentially modified, which gradually goes beyond the limits of "Uniterm" type.

In order to consider cases analogous to the above-mentioned example there is no need to reject the basic idea according to which the contents of inquiries and documents is transmitted by their dictionary composition. This idea, however, must be supplemented by another, in accordance with which between terms essential for transmission of the contents of texts there can exist relationships which we call basic — relationships by virtue of which the inquiry and document can be relevant even without the entry of the dictionary composition of the inquiry into the dictionary composition of the document.

At present successfully exploited systems of the "Uniterm" type are well-known, and it can appear that this contradicts the conclusion that one of the assumptions on which these systems are based, are refuted by the practice of their exploitation. This, however, is not so.

The basic relationships can be considered in the process of functioning of retrieval systems by various methods. First of all it is possible to fix the necessary relationships between terms, let us say, having assigned these relationships by list, — and one way or another introduce them into the retrieval system, for example, having assigned an algorithm of comparison of inquiries and documents using these relationships. Thus, for example, wishing to look up the above-mentioned example, it would have been possible to introduce into the set of terms the asymmetrical relationship of "subordination" and to subordinate the term "repair" to the term "exploitation," the term "circuit-breaker" to the term "equipment," and the term "small-oil" to the term "high-voltage" and to say that the document should be issued in answer to the inquiry only if every term of the inquiry either enters the document itself or is represented in it by a term subordinated to it.

If the problem is to construct an automatic retrieval system or at least a system which would not use the creative abilities of the person exploiting it in the process of functioning, then this method of realization of the basic relationships is the only method. Resorting to this method, we rise to a way which essentially changes the structure of the reorganized exploration system and through a number of intermediate systems leads to systems represented at present by "Pusto-Nepusto" systems. If, however, it is not necessary to limit human participation in the retrieval process, then the basic relationships can be realized even without changes in the retrieval system, with only one complication of the procedure of its use. The additional burden connected with realization of the basic relationships is borne not by the retrieval system but by the person exploiting it.

One of the variants of such a complication consists in the fact that in the retrieval pattern of the document there are inscribed not only terms present in it but also those the presence of which in inquiry would require, in the opinion of the indexer, delivery of the indexed document. Thus, for example, in the example examined by us, in the retrieval pattern of the document there should be included not only the terms "repair,' "small-oil," and "switch" but also the terms "exploitation," "high-voltage," "equipment" and, possibly, still others at the discretion of the indexer.

Another, in many respects stronger, variant of complication of the diagram of exploitation of the retrieval system -- of a complication also having the purpose of realization of basic ratios, consists in the fact that instead of one retrieval on one inquiry there are conducted a number of retrievals on several inquiries, which are modifications of the initial inquiry. Thus, in the example considered by us the inquiry "exploitation of high-voltage equipment" could have been possible at the discretion of the inquirer supplemented by such of its modifications as, let us say, "repair of high-voltage equipment," "exploitation of small-oil circuit-breakers," "repair of small-oil circuit-breakers," etc., and retrieval could have been carried out according to each of these inquiries.

Thus, using the "Uniterm" system, which issues a document only if its retrieval pattern contains all terms entering the retrieval pattern of the inquiry, we can by special procedures of exploitation of this system find in the end those documents which correspond to our inquiry, not including its dictionary composition. In all known cases of the successful use of exploration of "Uniterm"-type systems there are used both above-described methods of their complicated exploitation, and the need for such special measures directed towards removal of undesirable consequences of the third assumption indicates that namely this assumption was not justified.

As has already been said, in those cases in which the basic relationships have to be realized not by the retrieval system itself but by a scheme of its use, the developer of the retrieval system cannot make the basic relationships the object of special development, thereby shifting the burden and responsibility to indexers and inquirers. But if for considerations, such as the requirement of complete automaticity the realization of basic relationships is assumed by the retrieval system, special development of a system of basic relationships turns out to be inevitable.

During development of the basic relationships it is possible to lean on the most diverse assumptions, which can be confirmed or refuted only by experiment. Therefore, it is natural to start from attempts to solve the problem by the simplest means. During development of the "Pusto-Nepusto-4" as such means there were used, first, a retrieval language, analogous to the langauges of "Uniterm"-type systems the words of which – descriptors – were with rare exceptions translations of natural language terms, and, secondly, a transitive, asymmetrical and unreflexive predicate partially regulating the set of descriptors. The system of basic relationships was constructed as a set of sentences of type [P(d, d')] (П(д, д')) assigned by list, where d and d' are descriptors of the language, and P is a regulating predicate.

Thus, one of the assumptions on which the logic of "Pusto-Nepusto"-class is based, consisted in the possibility of reaching the necessary result with help of paired ratios of the form P(d, d') not depending on context where P is a predicate of partial order. Both the operational experience of the "Pusto-Nepusto-4" system and the first series of experimental retrievals via the "Pusto-Nepusto-2" system do not yet give sufficient bases for refutal of this assumption, although it is already clear that certain advantages deserving consideration could be given by a system of relationships depending on the context of the document and inquiry compared. Such dependence can be realized in many ways – by both direct and roundabout methods detailed discussion of which is not within the scope of the present report. Let us note only that fixing of word combinations and homonyms introduced in insignificant quantity into the dictionary of "Pusto-Nepusto" systems as independent descriptors permits making the necessary basic relationships to a certain extent essentially dependent on the context of the documents. However, the number of word combinations and homonyms introduced into the language is small, so that we do not yet have bases to talk about systematic use of connections depending on context.

Thus, the basic relationships are introduced specially so that a document pithily corresponding to the inquiry could be issued also when it does not include the whole dictionary composition of this inquiry. Therefore, the simplest principle of setting basic relationships P(d, d') between descriptors d and d' taken in this order is the following principle: let us assume that [D] (Д) and D' are random documents, and let us assume that the descriptor pattern of document D' is obtained from the descriptor pattern of document D by replacement of descriptor d with descriptor d'; if document D' is relevant to any inquiry relevant to document D, between descriptors d and d' there should be established relationship P(d, d').

It is probable that certain pithy relationships between ideas can be intimately connected with the formally determined relationship P. Thus, for example, it is possible to expect that generic relationships will be included in relationship P in the sense that every time d is a generic idea with respect to form d', it will be necessary to establish relationship P(d, d') between them. On this count it is possible to express many assumptions, which, however, all need experimental check. It would be interesting, for example, to clarify whether the relationship of type to form corresponds with that part of relationship P which does not need establishment of dependence on context. However, it is important to emphasize that the connection of the formal relationship P with pithy relationships is an empirical fact and cannot be obtained as a result of a deductive conclusion. In this connection it is interesting to note that such an approach to the establishment of basic relationships was clearly formulated, as far as we know, only in two cases: by the Nidkhema group in Great Britain – as we learned of this in 1963 during the [FID] (ФИД) [expansion unknown] symposium in Moscow and during construction of systems of the "Pusto-Nepusto" class.

Appearance in the logic of retrieval systems of fixed basic relationships requires making new decisions which need to be checked experimentally and can be initially based only on a priori assumptions.

First of all one should note that the above-formulated principle of establishing relationship P(d, d') does not give us any algorithm which would indeed allow deciding whether or not relationship P should be established between random d and d'. This principle can therefore be considered only heuristic help of our intuition, and only exploitation of the retrieval system can show, how successfully this principle can be put into practice.

Further, even if it is assumed that the principle of establishing relationship P can be conducted in some sense in series and sufficiently effectively, then during formulation of comparison rules nevertheless it is necessary to make a whole row of assumptions about this relationship.

For convenience of the following presentation we will say that descriptor d is subordinate to itself descriptor d' or stands higher than this descriptor if relationship P is established between them. Then assumptions made during the construction of the "Pusto-Nepusto" system can be formulated in the following way.

The first assumption is that the relevance of the document is influenced not only by replacement of one of its descriptors with the descriptor directly below it, but also by simultaneous replacement of an arbitrary number of descriptors with arbitrarily descriptors below them. This assumption is not evident and does not come from the principle of establishment of basic relationships. Nonetheless one may assume that the practice of experimental exploitation of the

"Pusto-Nepusto" system confirmed this assumption for those concrete conditions in which the system was tested. It remains unclarified what role in confirmation of this assumption is played by the circumstance that almost half of the descriptors of language were not connected with any other descriptors; average length of circuits of interconnected descriptors did not exceed 3, and the average number of descriptors in an inquiry was equal to 4 (i.e., that during solution of the problem of document, replacement of not more than four descriptors was actual).

The second assumption used in the formulation of rules of comparison in the "Pusto-Nepusto-4" system is connected with replacement of descriptors with higher descriptors. It was assumed that relationship P should to a considerable extent coincide with the relationship of the general to the particular idea and that, consequently, replacing higher descriptors, we will probably make the subject of consideration more general. Therefore, in the rules of comparison it was anticipated that replacement of descriptors with higher ones lowers relevance but does not make it equal to zero.

On the same basis it was, further assumed that presence in the document of descriptros above the descriptors of the inquiry somewhat lowers document relevance since it is probable that the document is not about the subject of interest to the inquirer but about something more general.

The last two assumptions turned out to be not as well-founded. They were not confirmed by practice of exploitation of the "Pusto-Nepusto-4" system, and this circumstance was one of the causes of transition to the "Pusto-Nepusto-2" system.

The first of these assumptions lead to noticeable noise. This is explained apparently by the fact that on the one hand, relationship P is connected with generic relationships not as closely as it seemed to be at first, and on the other by the fact that transivity of relationship P led to delivery in answer to very concrete inquiries of a large number of such general documents that they must have been of no real value to the inquirer.

The last of the assumptions now examined could not lead to serious troubles, since it influenced not the composition of issued documents but only their distribution with respect to the echelons of delivery. Nonetheless the unnaturalness of distribution of documents by echelons was in certain cases so evident that this had to be grasped by the inquirer as a deficiency.

The following assumptions were made in connection with what has been said on the basis of the "Pusto-Nepusto-2" system.

First of all, the "Pusto-Nepusto-2" system rests on that fundamental idea of Taub according to which the contents of the document and inquiry is transmitted by their dictionary composition with fullness sufficient for the purpose of information retrieval.

Further, as in the "Pusto-Nepusto-4" system it was assumed that for delivery of a certain document it is necessary and sufficient that every descriptor of the inquiry be represented in the retrieval pattern of the document either by a descriptor equal to it or by some descriptor connected with it by basic relationships.

But in contrast to the "Pusto-Nepusto-4" system it was now necessary to reject the assumption that the necessary result can be reached with the help of *one* transitive predicate P, partially regulating the set of descriptors. Instead of this the basis of logic of the "Pusto-Nepusto-2" was the assumption of the *two* predicates P^1 and P^2. The first of them coincides by and large with predicate P of the "Pusto-Nepusto-4" system and P^1(d, d') can be as before read "d' is below d." The second predicate in contrast to the first is *not transitive* and to a well-known degree can be grasped as a predicate having a *large number of exceptions* reverse to P^1: in a noticeable number of cases P^1(d, d') is equivalent to P^2(d, d'). P^2(d, d') is read "d is above d'." Thus, in the "Pusto-Nepusto-2" system "d is above d'" and "d' is below d" — this is not one and the same, and, furthermore, it can happen that "d is above d'," "d' is above d'," but "d is not above d'."

The basic relationships of the "Pusto-Nepusto-2" system were developed according to the principle usual for "Pusto-Nepusto" systems which is described above and which was used in its time by Nidkhem. They remained independent of context just as in the "Pusto-Nepusto-4" system and with the same reservations with respect to word combinations and homonyms.

It is necessary to note that the generalization of basic relationships made in the "Pusto-Nepusto-2" system as compared to the "Pusto-Nepusto-4" system is not, of course, the only conceivable one. But it is the simplest one which allows counting on removal of deficiencies revealed in the logic of the "Pusto-Nepusto-4" system.

The rules of comparison of the "Pusto-Nepusto-2" system also rest on assumptions different from the corresponding assumptions of the "Pusto-Nepusto-4" system. These assumptions can be formulated in the following way:

1. If an inquiry descriptor in a document is replaced with a lower descriptor, this in no way reflects on the relevance of the document.

2. If a document for a certain inquiry descriptor has not only an equal or lower but also a higher one, then this in no way effects document relevance.

3. If in the document for a certain inquiry descriptor there is neither an equal nor a lower one, but at least one higher one, then this lowers document relevance, but does not make it equal to zero.

From what has been said it is easy to see that "Pusto-Nepusto-2" logic makes it possible to split delivery only into 2 echelons against the 4 echelons of the "Pusto-Nepusto-4" system, where echelons of the "Pusto-Nepusto-2" system cannot be obtained by grouping echelons of the "Pusto-Nepusto-4" system.

Now the "Pusto-Nepusto-2" system is in experimental exploitation and at the present it is early to draw conclusions about results of conducted reorganization of the logic of systems of the "Pusto-Nepusto" class. Nevertheless, preliminary data bear witness to the fact that delivery of the new system is fuller than that of the old one and contains noticeably less noise. For the present it is difficult to say how essential the loss of the possibility to divide delivery into four echelons is: if this proves to be an essential deficiency of the new system, then it will most likely be essential only in arrays of the order of several hundred thousand, which we have not yet achieved.

THE "SETKA-3" AUTOMATED IPS ON THE "MINSK-22" WITH THE USE OF THE SOCKET ASSOCIATIVE-ADDRESS METHOD OF ORGANIZATION OF INFORMATION

S. A. Gorokhov

Since 1964 at [NIIEIR] (НИИЭИР) [expansion unknown] there have been developed several fundamentally different approaches to construction (machine realization) of automated information-retrieval systems ([IPS] (ИПС)) of the descriptor type on the "Minsk-22" and the "Minsk-2" [1]. There were examined IPS using in diverse variants the nodal associative-address method of organization of initial information and the "zonal" method with various principles of coding initial information. In spite of the fact that during the associative-address method of retrieval there is accepted direct organization of initial information, it turned out to be possible for such IPS to use the inverse method of retrieval [2], which permitted significantly lowering the consumption of machine time expending realizing inquiries.

As criterion of semantic conformity of the contents of the inquiry to the contents of the document in the mentioned IPS there is accepted the entry, sometimes with certain elements of grammar, of all the descriptors of the inquiry into the retrieval pattern of the document [POD] (ПОД).[1] In spite of its logical simplicity,

[1] Under elements of grammar here there is understood appropriation of weight (0 or 1) to descriptors of inquiry depending upon the semantic load which they carry in the inquiry, and the connection of descriptors in the inquiry by the clusters "AND" and "NOT."

which made it possible considerably to simplify the IPS algorithm, the given criterion of semantic conformity together with principles laid during indexing of the document ensures sufficiently high output characteristics of IPS.

Below there is examined an improved variant of the "Setka-3" IPS on the "Minsk-22" with use of the socket associative-address method of organization of information.

As initial information in the examined IPS there are taken two thematic divisions from [GSBK] (ГСБК) [expansion unknown] NIIEIR — "Computer technology," consisting of two parts (20 thousand and 12 thousand documents), and "Transformers" (1.3 thousand documents).

I. Basic Characteristics of the "Minsk-22"

For the best understanding of certain sides of the work of IPS essentially connected with ETsVM possibilities, we will give the basic characteristics of the "Minsk-22."

The "Minsk-22" is a two-address ETsVM with a speed of 5-6 thousand operations per second. Its calculation grid consists of 37 bits. Magnetic working storage (internal memory) consists of ferrite cores and contains 8192 37-bit words. External memory is magnetic-tape storage consisting of 16 tape-drive mechanisms. On any tape-drive mechanism there can be set magnetic tape accomodating an average of 75 thousand 37-bit words. On magnetic tape information is recorded in zones. The volume of one zone is 2048 words. Information is recorded on magnetic tape from fast store in any place of a zone, and readout of information from magnetic tape to fast store is also possible from any place of a zone. It is possible to put out 4096 words in two zones in succession.

The "Minsk-22" allows input of numerical (binary and binary-decimal system) and alphabetic (Cyrillic and Latin alphabets) information. Initial information can be fed to the "Minsk-22" both from punched tape and from punched cards.

Calculation results (numerical and alphabetic information) are derived differently. Numerical information in octal and decimal systems is printed at a rate of 20 twelve-character lines per second.

Alphabetic information is printed on an alphameric printer ([ATsPU] (АЦПУ)). The maximum value of a line of ATsPU printing is 128 symbols, and printing speed is 7 lines per second.

II. Information-Retrieval System

Application of ETsVM for information retrieval essentially constitutes an attempt to ensure more convenient and rapid access to accumulated knowledge, so that developers and scientific workers are given timely and complete scientific and technical information needed by them in their work.

Examining the work of automated IPS, it is necessary to note that in an absolute majority of contemporary IPS ETsVM "take over" the main part of the "mechanical work": retrieval in a large array of scientific and technical information but one built in a certain way, delivery of answers to inquiries in a predetermined form, etc. The person servicing the IPS in this case indexes documents for putting into the IPS, which is connected with semantic appraisal of the contents of documents of scientific and technical information, semantically analyzes and indexes entering inquiries, composes a dictionary of descriptors, etc. It is obvious that in the near future some of these functions will be wholly and some partially fulfilled by ETsVM. Certain prerequisites to this will be shown below.

Let us consider now the construction and functioning of the main parts of automated IPS developed at the NIIEIR.

1. Creation of a Dictionary of Descriptors for the Thematic Division "Computer Technology"

The first stage in the work of creating an automated-IPS language was the indexing of documents of the thematic division "Computer technology." The indexer was tasked with as much more exactly and briefly as possible expressing the basic contents of the indexed document. The indexer need not limit himself to the terminology used in a given document, when it is necessary to use a word more exactly or more generally copying the contents of the document, although not contained in it.

Thus, on the first stage of work there was created a POD card file of the division "Computer technology." Division POD on the average consist of 8 words. The greater part of these words expresses the basic semantic contents of the document, and the lesser bibliographic data about the document and the name of the equipment or development on the whole in the given document.

The following stage was connected with the work of ETsVM and consisted in counting frequency of recurrence of various POD terms. For this the words in the POD were normalized, that is, they were reduced to the standard form of recording – masculine gender, nominative case, and singular number. Stable word combinations of the type "arithmetic unit," "magnetic drum" and similar ones were replaced by the abbreviations – ["au," "mb"] ("ау," "мб"). From the POD there were excluded prepositions, conjunctions, and other words not essential for transmission of the basic contents of documents. Normalized POD obtained after that were perforated and were introduced into the ETsVM. Then a special program calculated frequency of recurrence of every word in the array of normalized POD and these data were printed.

The total number of normalized words in the POD of the first part of the division "Computer technology" exceeded 120 thousand. Sixty-five hundred different words were obtained. The frequency of recurrence of individual words varies from one to 2000.

The next stage of work is composition of the dictionary of descriptors. By the table of frequency of recurrence of words in the POD there were removed words the frequency of recurrence of which is highe· than a certain numerical threshold (in the examined case the numbe. 10). The total number of these words was 600-700.

These words formed the main part of the dictionary of descriptors. From a large part of the words the frequency of recurrence of which is lower than the threshold shown, there were formed classes of equivalence. The words of the remaining smaller part were either replaced in the POD with words which are encountered more frequently and are then introduced into certain classes of equivalence or removed from consideration.

In a class of equivalence there were united words the presence of one of which in the inquiry with a high probability will require delivery of documents, indexed in other words of the given class of equivalence.

One of the words of the class of equivalence, usually the one most fully expressing the semantic value of the given class was called a descriptor, and all the remaining words of the class of equivalence were called key words. Along with frequently encountered words to descriptors there was advanced a certain part of the words perspective for the field of computer technology but encountered fewer times in the array.

Thus there were formed 814 classes of equivalence, that is, 814 descriptors were determined.

Words which are close in meaning, but not close enough to be united into one class of equivalence, are supplied with the reference "see also." This reference is needed later for possible expansion of delivery of answers to inquiries.

Besides semantic descriptors, in the dictionary there were included 34 descriptors of a bibliographical character and a certain number of descriptors designating the names of equipment, developments, and firms.

Thus, in the dictionary of the thematic array "Computer technology" there were included about a thousand descriptors.

2. Construction of a Machine Dictionary of Descriptors of an Automated IPS

The carrier of the machine dictionary of descriptors of the improved variant of the "Setka-3" automated IPS on the "Minsk-22" is punched cards.

In the upper part of the punched card with the help of a typewriter there is recorded the designation of the descriptor or key word and reference number in the dictionary of descriptors of the class of equivalence to which the given key word belongs or which the given descriptor determines.

Then two lines are punched in the punched card. In the first of them there is shown the first information line ([IS_1] ($ИC_1$)) of the descriptor subarray of the given class of equivalence, and in the second the second information line (IS_2). Their structure and assignment will be pulled apart below. The descriptor and all key words entering the class of equivalence of the given descriptor have identical IS_1 and IS_2.

All punched cards of the machine dictionary of descriptors are collected in a card file, in which they are located in alphabetic order of descriptors and key words. On the average the card file contains every punched card in triplicate or quadruplicate, since answers to several inquiries can be retrieved simultaneously in the IPS under consideration. Furthermore, there are a certain number of punched cards with the designation of operation of negation.

3. Processing of Inquiries and Their Input into ETsVM

Inquiries for retrieval with the help of automated IPS initially have no limitations placed on them, besides the wish to most exactly formulate the object of retrieval.

Then the specialist servicing the IPS according to the given thematic division analyzes the incoming inquiry and indexes it (that is, translates the contents of the inquiry into the terms of the dictionary of descriptors of the thematic division); after that it is determined whether the inquiry is complicated or simple. A simple inquiry is one which belongs to only one thematic division and the descriptors in which are united by clusters "AND" and "NOT." Any inquiry not satisfying these conditions is complicated. A complicated inquiry is reduced to the sum of simple inquiries if possible, or is reformulated.

Then punched cards with the descriptors of the simple inquiry will be selected from the card file of the machine dictionary of descriptors. The punched card of the negated descriptor is proceeded by the punched card of negation. Inquiry is separated from inquiry for machine realization by a special punched card.

The group of inquiries thus selected (the maximum number of inquiries in the group must not be over 36) is put into the reader and fed into working storage via programming.

4. Recording of Initial Information in ETsVM

Initial information for retrieval in automated IPS is an array of reference numbers of documents in GSEK NIIEIR according to the corresponding thematic division of documents.

In the examined IPS there is accepted the inverse method of retrieval and -- accordingly -- the inverse form of organization of initial information, i.e., the array of initial information is recorded in the form of so-called descriptor subarrays.

Descriptor subarrays are sets of reference numbers of documents recorded in order of increasing absolute values and pertaining to one descriptor (class of equivalence). The number of these subarrays is determined by the number of descriptors in the dictionary, and their dimensions by the frequency of occurrence of the given descriptor in the POD of the examined thematic division of documents.

For the purpose of saving memory and hastening retrieval in IPS there is accepted position coding of descriptor subarrays and the two-stage method of retrieval of documents[1] answering the inquiry. The principle of position coding consists in the fact that the number of a document is not written in the form of a number in this or that number system, but is determined when necessary as the number of a position reckoned from a certain origin. A working storage bit was taken as a position element. Since one or zero can be recorded in any bit of a working-storage cell, then let us agree to mark an occupied position one, and a free one zero.

Thus, a working-storage cell can be considered a 36-position section, which will henceforth be called a position interval. However, during direct position coding of an array of initial information descriptor subarrays become sufficiently large in value and a large part of the position intervals in them turn out to be empty. Therefore, indirect position coding was used because position-interval structure have to be modified.

Every position interval was ascribed its reference number in the general sequence of position intervals of the examined thematic division. The position interval was cut down to 25 bits, and the remaining 11 bits were for recording the reference number of the position interval. All empty position intervals were removed from the descriptor subarrays.

The subarrays of numbers of documents consisting of position intervals not empty belonging to a certain descriptor is called the second descriptor subarray and the position intervals of this subarray are called second position intervals.

To speed up retrieval every descriptor from the dictionary was compared with a subarray — the subarray in which direct position code

[1] At present the possibility of a three-step retrieval method is under consideration.

marked all numbers of position intervals entering the second descriptor subarray of the given descriptor. This subarray is called the first descriptor subarray, and position intervals of this subarray are called first position intervals.

First descriptor subarrays have constant value for all descriptors of one thematic division of documents. Their value is determined by the total number of documents in this division.

The first and second subarrays in the ETsVM are recorded on magnetic tape in two large groups. In first group there are extracted all first descriptor subarrays and into the second all second ones.

So that any of the first descriptor subarrays can be later supplemented each of them ends in a set of empty cells. On magnetic tape the first descriptor subarrays are recorded in order of decreasing number of position intervals in the corresponding second descriptor subarrays.

Second descriptor subarrays are subarrays of variable dimensions. Position intervals in them are located in order of their increasing numbers. On magnetic tape one subarray is separated from another by a set of empty cells. In the last cell of this set after complete filling of it there is placed the address indicating the place of recording on magnetic tape and the dimensions of the new subarray, which is a continuation of the completely filled set. Thus the recording of second descriptor subarrays is turned into a socket associative-address structure.

On magnetic tape second descriptor subarrays are recorded in order of decreasing number of numbers of documents in them.

Let us consider now the structure of the first and second information lines.

Each of the subarrays of any descriptor is set in conformity with an information line fixed in the dictionary of descriptors. It indicates the ·lace of the given descriptor of the subarray on magnetic ·ape and its value.

The structure of the first information line follows:

$$№_{з} \quad A_{н} \quad n,$$

where $№_{з}$ is the number of the magnetic tape zone on which the given first descriptor subarray is recorded; $A_{н}$ is the initial address of the first descriptor subarray in the zone; and n is the number of cells occupied by this subarray.

It is assumed that all first descriptor subarrays are recorded on magnetic tape hung on the zero tape-drive mechanism.

The structure of the second information line is

$$№_{лпм} \quad №_{з} \quad A_{н} \quad m,$$

where $№_{лпм}$ is the number of the tape-drive mechanism on which the corresponding magnetic tape is hung; $№_{з}$ is the number of the magnetic-tape zone, on which the second descriptor subarray is recorded; $A_{н}$ is the initial address of the second descriptor subarray in the zone; m is the number of cells occupied by this subarray.

The number of the tape-drive mechanism is shown because second descriptor subarrays can be disposed on several magnetic tapes hung on different tape-drive mechanisms accessible simultaneously.

5. Description of the Algorithm of Work of the IPS

Documents answering an inquiry in the examined IPS, are found in two stages.

The first stage is the stage of rough retrieval. On this stage there are selected numbers of second position intervals common to all the descriptors of the inquiry. As a result of this there is considerably narrowed the scope of information coming into play in the subsequent processing.

The second stage is the stage of sampling of common position intervals in the second descriptor subarrays of the inquiry and separating in them of the numbers of documents common to all the descriptors of the inquiry, which by virtue of the accepted criterion of semantic conformity signifies finding the numbers of documents answering the inquiry submitted.

Introduction of the two-stage method of retrieval permitted accelerating the retrieval process and made possible simultaneous retrievals in response to 36 inquiries.

Let us consider in detail fulfillment of each of these stages.

A group of information lines of inquiry descriptors is put into working storage from punched cards. After that the group is split up into two subarrays – the subarray of the first information lines and the subarray of the second information lines and table T_1 is formed, in which there are fixed the number of the inquiry, the number of descriptors (information lines) in it, and the initial address of the subarray of the first and second information lines of the corresponding inquiry. Information lines of negated descriptors are recorded with minus signs.

After that among positive IS_1 [information lines] there are those the number of the magnetic-tape zone of which is maximum and minimum. The zone with the minimum number is read into fast store and from it there are separated the necessary first descriptor subarrays, which are then put with the help of the operation of logical multiplication into an earlier-prepared place in fast store. Preparation of the place in fast store consists in recording ones in all bits of a certain set of successive cells, which will allow carrying out logical multiplication even for the first subarray.

After full processing of the zone its number is compared with the maximum number found. If these numbers did not match, a new minimum zone number is found among the remaining positive IS_1 and the zone found is processed as described above. If, however, the zone numbers matched, rough retrieval is over with and it is necessary to go to a new stage of processing – checking the equality to zero of all the lines of the result of processing of the first descriptor subarray of each of the inquiries. In case these lines are equal to zero the given inquiry is not further processed. Its number is printed with a minus sign. Otherwise, after completely checking the whole array, one goes to the next stage of processing.

The next stage of processing is preparation of a place in fast store for the second part of retrieval – finding the numbers of documents answering the inquiry.

There is formed table T_2, in which there is noted number of inquiry, number of second position intervals which must be further processed in response to the given inquiry, and the initial address in fast store of the corresponding subarray of second position intervals.

A place in fast store for the second stage of retrieval is prepared in the following way. In the first 25 bits of the cells there are stored ones, and in the 11 last ones the number of the second position interval for which further finishing is required.

Thus, for each of the inquiries there are recorded as many lines, as there are position intervals in it requiring finishing.

The second part of retrieval starts with finding the maximum number of tape-drive mechanism and the maximum zone on it and also the minimum number of tape-drive mechanism and minimum zone on it among second information lines.

The minimum zone is read into fast store, and from the corresponding second descriptor subarray there are selected second position intervals which require finishing, and with the help of logical multiplication they are put into their place in fast store. The position part of the second position intervals of negated descriptors before superposition is inverted with the help of logical multiplication.

Second position intervals are processed similarly until the maximum zone on the maximum tape-drive mechanism has been reached.

The last processing stage is the decoding of the position code and delivery of answers to inquiries. Position code is decoded by the formula:

$$\text{№}_{\text{пи}} \cdot 25 + P,$$

where $\text{№}_{\text{пи}}$ is the number of the position interval and P is the number of the position in the given position interval.

6. Forms of Deliveries of Answers to Inquiries

In the examined IPS provision is made for two forms of deliveries of answers to inquiries.

The first form is delivery of reference numbers of documents from the corresponding thematic division GSBK NIIEIR.

The second form is delivery of bibliographic descriptions of documents recorded under the given numbers. Delivery of bibliographic descriptions of documents is carried out at the option of the inquirer.

Examples of deliveries follow:

```
+ 0000 0000 0001
+ 0000 23951
+ 0000 0000 0002
+ 0000 23953
```

where the first and third lines designate the reference numbers of inquiries, and the second and fourth the reference numbers of abstracts issued in response to the corresponding inquiries.

Or:

Inquiry 1.

23951. Programs of addition and subtraction of sine with floating point on the "Ural-1." Klokachev I. V. In the collection "Solution of engineering problems with electronic computers." L., 1963, 8-13.

Inquiry 2.

23953. (Algorithms in Algol-60). Wegstein I. H. Algorithms "Communs Assoc. Comput. Mash." 1963, 6, No. 8, 441-450 (English).

7. IPS Characteristics

Let us give some IPS characteristics.

1. Number of inquiries simultaneously serviced.................................. 36
2. Average time (machine) of retrieval for one inquiry................................ 5-7 s
3. Number of magnetic-tape zones necessary for storage:
    - a) of descriptor subarrays (20 thousand documents)...............................
    - b) of bibliographic descriptions (20 thousand documents)..................... 300-400
4. Number of inquiries satisfied per shift:
    - a) answer in the form of reference numbers of documents..................... 1-1.5 thousand
    - b) answer in the form of bibliographic descriptions of documents............... 0.5 thousand

Bibliography

1. Bud'ko N. S., Kudin N. I., Gorokhov S. A. Avtomatizirovannaya informatsionno-poiskovaya sistema dlya istochnikov nauchno-tekhnicheskoy informatsii po radioelektronike "Setka-3" (The "Setka-3" automated information-retrieval system for sources of scientific and technical information on radio electronics), NTI, 1966, No. 10, 15-25.

2. Vleduts G. E., Lakhuti D. G., Chernyavskiy V. S. Ob inversnom printsipe realizatsii informatsionno-poiskovykh sistem (The inverse principle of realization of information-retrieval systems), NTI, 1963, No. 4, 37-41.

3. Mikhaylov A. I., Chernyy A. I., Gilyarevskiy R. S. Osnovy nauchnoy informatsii (Bases of scientific information). Izd-vo "Nauka", 1965, 247-339.

4. Ledli R. S. Programmirovaniye i ispol'zovaniye tsifrovykh vychislitel'nykh mashin (Programming and use of digital computers). Izd-vo "Mir", 1966, 583-626.

EXPERIENCE IN CREATING INFORMATION-RETRIEVAL LANGUAGE VIA COMPUTER TECHNOLOGY

V. K. Vakhabov, A. A. Mikhaylova, L. M. Yesilevskaya and
T. S. Kutayeva

At the Perm Scientific Research Institute of Control Machines and Systems attempts are being made to create an automated information-retrieval system [(IPS)] ((ИПС)) for a reference and information fund [(SIF)] ((СИФ)) of the [ONTI] (ОНТИ) [Association of Scientific and Technical Publishing Houses] of the instrument.

An important IPS element is information-retrieval language [(IPYa)] ((ИПЯ)). The present report reports on the development of information-retrieval language of the descriptor type according to the division of "Computer technology."

During selection of IPYa structure there were considered the following peculiarities of IPS operation.

1. High productivity of retrieval (up to three-four thousand inquiries in a day) via application of a magnetic-drum electronic digital computer.

2. The need for machine translation of key words in the retrieval instruction into codes of descriptors to increase retrieval productivity.

3. Absence of direct feedback with the user during machine retrieval for correction of the retrieval instruction for the purpose of obtaining the required fullness and accuracy. Feedback is achieved in the system via application of a three-circuit IPS system, where the first circuit does not have feedback (Fig. 1).

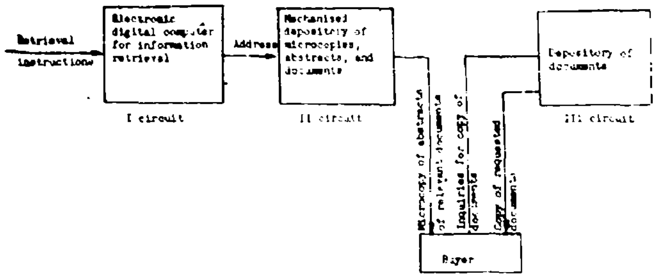

Retrieval instructions
Electronic digital computer for information retrieval
Address
Mechanised depository of microcopies, abstracts, and documents
Depository of documents
I circuit
II circuit
III circuit
Microcopy of abstracts of relevant documents
Inquiries for copy of documents
Copy of requested documents
Buyer

Fig. 1. Block diagram of three-circuit IPS.

Under such conditions presence of noise requires only a certain increase in the productivity of the secondary circuit. Therefore, noise is less important than losses.

4. The whole information array in the branch center is split into a number of big thematic subarrays with a volume of the order of 30 thousand documents. For each of the subarrays its own local IPYa is developed.

The enumerated peculiarities of IPS operation determine the most important features of the information-retrieval language developed.

1. The language has basic relationships of the type "higher — lower" between ideas in order to ensure delivery of documents concerning particular ideas on an inquiry formulated in more general ideas, and, thus, lower losses.

2. It was decided to introduce grammatical means very carefully, only after experimental measurements of noise. At present it has been decided not to intorduce grammar. If noise with increase or array exceeds 50%, then in the first place one should apparently

introduce grammar of the type of indicators of communication. Application of role indicators is problematic since available source materials [2], [3] show that use of role indicators leads to high subjectivity of indexing and does not lower noise much when losses increase considerably.

3. The criterion of semantic conformity of the developed language is simple — on entry. Basic relationships are considered during indexing in the following way: if the descriptor, included in the retrieval pattern has dispatch to a higher descriptor, then the indexer includes the higher descriptor in the retrieval pattern. This is equivalent to an insignificant increase in depth of indexing but does not complicate the criterion of semantic conformity. The simplicity of the criterion of semantic conformity is the condition of high productivity of retrieval.

4. Depth of indexing averages 8-10 descriptors per document. As results of experiments of Soviet and foreign specialists show [1] and [4], increasing the number of descriptors in the retrieval pattern markedly increases noise when fullness increases insignificantly.

5. Presence of a machine dictionary for translation of the key words of an inquiry into codes of descriptors inevitably requires a certain standardization of key words of the dictionary and retrieval instruction.

The following basic rules are accepted:

a) a majority of key words are separate words of natural language. Word combination is used only in the case when it is a commonly used scientific term. It can correspond to the abbreviation, which is also included in the dictionary. For example: computer — [VM] (ВМ), memory unit — [ZU] (ЗУ);

b) key words have to be nouns, adjectives, rarely numerals;

c) all words are singular, with the exception of those words having no singular;

d) adjectives are masculine. The dictionary was composed on an array of 1300 abstracts on computer technology from [RET] (PЭT) [expansion unknown] journals.

In the process of free indexing of abstracts there were selected key words, they were integrated into classes of conditional equivalence, and basic links were established in the form of references "see" to higher descriptors. Upon the termination of this work the dictionary contained 664 words and 367 descriptors (classes of conditional equivalence). Then on the basis of the available dictionary 1060 abstracts were indexed. New words were added to the dictionary. At present the dictionary contains 702 key words in 404 classes of conditional equivalence.

From these data it is possible to trace the character of dependence of the value of the dictionary on the volume of the information array (Fig. 2). From the given graph it is clear that growth of the dictionary is considerably delayed when the array of documents increases. This phenomenon is called dictionary saturation.

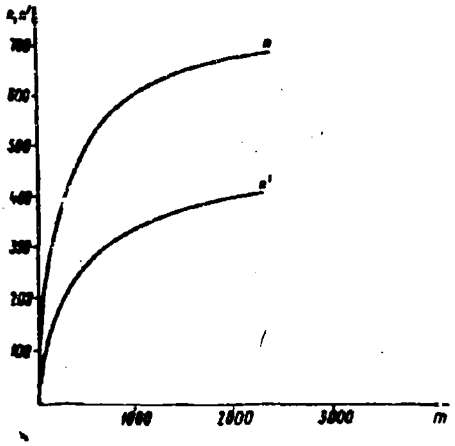


Fig. 2. Dependence of the volume of the descriptor dictionary and the dictionary of key words on the number of documents in the array: m is the number of documents in the array, n' is the number of desciiptors in the dictionary, n is the number of key words in the dictionary.

To evaluate the language developed there was conducted an experiment on an initial array of 1300 documents and on a summary array of 2360 documents. The purpose of the experiment was:

1) to determine noise factors and losses;

2) to clarify how these indices vary with increase in the retrieval array. This is necessary to forecast introduction of grammatical means into the IPYa.

For the experiment there were formulated 250 inquiries. They were composed by specialists not participating in IPYa development. In every inquiry there were no fewer than three key words, a maximum of nine, and an average of five key words in an inquiry. On the basis of these inquiries the coefficient of accuracy was calculated by the formula [2]:

$$A = \frac{100R}{L}\%,$$

where R is the number of relevant documents in the delivery, and L is the total number of documents in the delivery.

To calculate the coefficient of accuracy retrieval was carried out for 150 inquiries. For all 150 inquiries 412 documents were hits, of which 351 were relevant.

Analysis showed that for 113 inquiries only relevant documents were hits, and for 37 inquiries, besides relevant documents, documents not answering an inquiry were hits.

Of 37 inquiries 22 drew one unnecessary document each, 10 drew two, and the other five inquiries each drew 3 or more documents. After a study of causes of errors it turned out that 7% of information noise was caused by indexing deficiencies and 93% by irremovable noise through false combinations.

Example 1. Inquiry: principle of action of core storage. The total number of documents issued to the inquiry is six, five relevant. One document does not answer the inquiry and was issued as a result of false combinations. The document talks about the principle of action of thin-film storage and the method of selection of words with the help of ferrite cores.

Example 2. Inquiry: characteristics of magnetic storage. The total number of documents issued to the inquiry is 22, 21 relevant, one superfluous (not answering the inquiry) document is issued via false combinations. The document talks about characteristics of a military electronic miniature system with magnetic ZU [storage].

Coefficient of fullness was calculated by two methods: first, as a percentage of the number of relevant documents in the delivery to the total number of relevant documents in the retrieval array [2]; secondly, as the ratio of the number of found initial documents to 100 inquiries [2]. A source document is the document from which the inquiry is composed. There are 100 initial documents.

In view of the complexity and labor-consuming character of finding the total number of relevant documents in the retrieval array, coefficient of fullness was determined by the first method for 10 inquiries. Coefficients of fullness calculated by the two different methods give the same result (see table).

Table.

Parameters of IPS effectiveness	Information array on which the IPYa was created (1300), %	Summary array on which IPYa was worked out (2360), %
Coefficient of fullness (first method)	92	92
Coefficient of fullness (second method)	92	92
Coefficient of accuracy	85	80

Losses of documents occur through the complexity of calculation of all aspects illuminated in the document during indexing.

Example. Inquiry: application of thermoregulators for normal ZU operation. Due to the absence in the retrieval pattern of the key word "Thermoregulator" the following text was not issued in answer to this:

Copy No. 9967	Universal Decimal Classification 681.142.652.2 Stabilization circuit of recording current of address circuit of type-Z ZU. "Information of inquired sheets." 1963 No. 3441, 3 p., illustrated.

There is described the circuit of stabilization of the address recording current intended for use in type-Z ZU containing 128 27-bit numbers consisting of (2 × 1, 4 × 0.9) [VT-1] (BT-1)-type ferrite cores. The cores operate under the following conditions: readout current $I_{сч}$ = -(1.2-1.5) A; discharge current of recording $I_{разр}$ = 0.6 A; address current of recording $I_{зап}$ = 0.7 A; fixed bias $I_{см}$ = 0.3 A. The stabilizing circuit consists of six [P25B] (П25Б) (or P25A) transistors and regardless of the number of reversed cores in the numerical rule (load) ensures stable current $I_{зап}$ = 0.7 A by way of limiting it by the internal resistance of the circuit. To ensure normal ZU operation when ambient temperature varies, in the stabilizing circuit there are used thermoregulators consisting of two semiconductor thermistors and a diode. The proposed circuit ensures normal ZU operation in the temperature range from -20 to +60°C. The results of the experiment are given in the table.

In the process of indexing the information array there was conducted a study of law of distribution of descriptors in the retrieval patterns of the documents.

As is known [5] and [6], the frequency of appearance of words of natural language follows the Zipf law with high accuracy

$$p_i = \frac{K}{i},$$

where K = const; i = 1, 2, ..., n is the reference number of word during location in order of decreasing frequency.

More exactly the distribution of the words of natural language is described by the Mandel'brot law, which essentially generalizes the Zipf law:

$$P_i = \frac{K}{(B+i)^{\alpha}},$$

where K, B and α are constants, where $1<\alpha<1,2$.

The study of the real law of distribution of descriptors in retrieval patterns of documents shows that Zipf and Mandel'brot laws known for distribution of words of natural language, also well describe distribution of descriptors.

Figure 3 shows the real law of distribution of descriptors. From the graph it is clear that this law can be described by the expression:

$$P_i = \frac{0.202}{2+i},$$

which fully agrees with the Zipf and Mandel'brot laws.

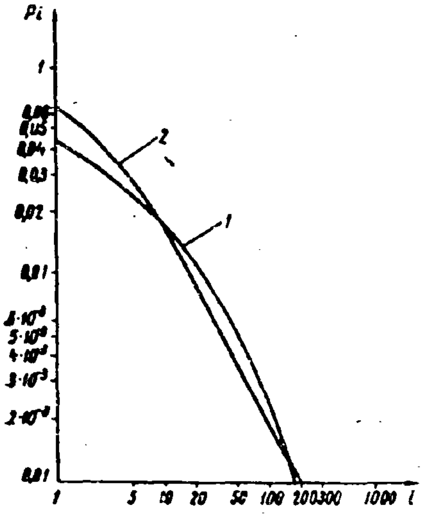


Fig. 3. 1 – the real law of distribution of descriptors in retrieval patterns of documents, 2 – curve plotted according to the law $P_i = \frac{0.202}{2+i}$.

Conclusions

1. Conducted experiments show that the developed IPYa, in spite of its simplicity of structure, has fully satisfactory characteristics.

2. Growth of dictionary is considerably delayed during the growth of an array of over 2000 documents.

3. Input of grammatical means into IPYa is inexpedient for small arrays. The question of needing to introduce grammatical means into developed IPYa will be examined after carrying out in 1967 experiments on an array of 15-20 thousand documents.

4. Distribution of descriptors in retrieval patterns of documents obeys the same Zipf and Mandel'brot laws, as words in natural-language texts.

Bibliography

1. Vickery B. Vocabularies for coordinate systems. "Aslib. Proc," 1963, 15, No. 6, 170-176.

2. Cleverdon C., Lancaster F., Mills I. Uncovering some facts of life in information retrieval. "Spec. Libr.," 1964, 55, No. 2, 86-91.

3. Lancaster F. W. Same observations on the performance of EJC role indicators in a mechanized retrieval system. "Spec. Libr.," 1964, 55, No. 10, 696-701.

4. Mikhaylov A. I., Chernyy A. I., Gilyarevskiy R. S. "Osnovy nauchnoy informatsii" ("Bases of scientific information"). M., Izd. "Nauka", 1965.

5. Brillyuen L. "Nauka i teoriya informatsii" (Science and information theory"). M., Fizmatgiz, 1960.

6. Mandel'brot B. O rekurrentnom kodirovanii, ogranichivayushchem vliyaniye pomekh (Recurrence coding limiting the interference effect). In the collection "Teoriya peredachi coobshcheniy". M. IL., 1957.

REALIZATION OF THEMATIC INFORMATION-RETRIEVAL SYSTEMS ON AN ELECTRONIC DIGITAL COMPUTER

M. P. Ubiyko

At present patent-licensing work in scientific research institutes is continuously expanding. Branch patent funds are growing; the volume of patent-description investigations is growing.

From patent funds there are determined the world technical level and direction of development with respect to one technical branch or another. With respect to the same funds there are made numerous examinations for the purpose of determining:

– the patent purity of articles or their components;

– the novelty of developments and inventions;

– the need to develop a new article or the expediency of obtaining licenses.

These operations take much time of specialists, inasmuch as it is necessary to examine hundreds and thousands of patent descriptions, in order to make a patent examination. Months are spent making an average examination. The main part of the time (over 50%) is taken retrieving patents.

Therefore, the tendency of many specialists, whole organizations, and institutes to simplify and accelerate these operations is natural. And apportionment of patents on narrow subjects is the first step in this direction.

The most essential shortening of the time spent retrieving patents is given by systematization of branch patent fund with help of well-known tabulyagrams [no translation found]. This method was developed by the Central Scientific Research Institute of Patent Information TsNIIPI.

Our organization from the very beginning of formation of the branch patent fund systematized it with the help of tabulyagrams.

Further operational experience showed that in such systematization, considerably facilitating and accelerating retrieval of patents, there is laid a real possibility for mechanization of retrieval with the help of ETsVM, which will reduce still more the time it takes to retrieve patents. All patent data in tabulyagrams are encoded by numerals.

In our organization it was possible to realize retrieval on the "Ural-2."

We mechanized such labor-consuming processes as:

— examination of all patents pertaining to the subject of interest with respect to the funds of one or several countries.

— sampling and recording of numbers of patent descriptions from patent funds of one country or a group of countries which pertain directly to the assigned subject.

— delivery of a typewritten reference of contents: = On the technical question interesting You in the country (in the countries) there are the following numbers of patent descriptions (there are reported the country and number of patents in the fund of this country). =

The only thing left for the specialist to do is to obtain patent descriptions on the issued reference in the patent library and to investigate them.

Thus, the specialist assigns the question and [illegible]tigates the patent descriptions found, and any other work is d[illegible] machine.

However, it is noticed that of the obtained and investigated patents the specialist finds only a few patents of interest to him, and sometimes nothing at all.

We now work on these, in order to mechanize this process, too, i.e., investigation. For this purpose, apparently, it will be necessary to train, figuratively speaking, a machine to answer more complicated questions.

For realization of the idea of computer retrieval of patents there was carried out a volume of works on programming of retrieval on the machine and the corresponding shift of the contents of the tabulyagrams to the storage units of the machine.

1. Volume of programming usually depends on the complexity of those tasks required of the machine and on the volume of operations which need to be carried out.

But since in this case the discussion concerns patent information, we are first of all interested in retrieving the patents we need from the whole volume of the branch fund without the participation of the specialist. Therefore, programming was faced with two simple questions:

– What numbers of patent descriptions are in one country "X" on the technical question interesting us (transmission system, form of propeller, and others)?

– What numbers of patent descriptions are in any group of countries on the same question or on any other technical question (switches, reduction gears, or transistors, and others)?

2. After that all patent information of one tabulyagram was split up into groups - so-called zones with appropriation to them of numbers. In accordance with the volume of the operational storage unit of the machine in each zone there should not be more than (4096 instructions) or 2048 numbers.

3. Then there was composed a so-called working table of operator with numbers of zones and addresses for input of information, in which there were the following columns:

1) numbers in order;

2) numbers of headings according to the tabulyagrams;

3) numbers of zones (recorded in the octal system);

4) address of memory cells for every zone (recorded in the octal system).

Initial addresses of all zones start with 1000. The final address is determined by the number of patents in the given zone. Part of the addresses - from 0 to 1000 - is assigned for telegram.

4. For input of initial information (patent information) of groups of numbers or a zone in the operational storage unit as information carrier there was selected standard opaque 35-mm film.

Such an information carrier is distinguished by its independence, which is very convenient in those cases in which an ETsVM is used mainly for other purposes. Furthermore, the volume of information in such a carrier is practically unlimited.

On film information from tabulyagrams is transferred by the method of punching on an external device of the machine (PFCh2).

The following data are recorded on punched tape:

- zone number;

- patent number;

- number of country to which patent belongs.

Numbers of patents are printed in the decimal system next to the number of the country also in the decimal system.

The patent number must not exceed an eight-digit number, and the number of the country a two-digit number. These conditions are easy to satisfy.

The data of the printing are thoroughly collated with the data of the tabulyagram and then the tape is glued in a circle, and it is ready for use. It is numbered correspondingly.

Punching of tape with transfer of information by one thousand patents is fulfilled in 3-4 shifts by one man. Data of 52 patents are accommodated in one meter of tape. The maximum permissible tape length is 250 m (about 12,000 patents). Input speed is 150-160 patents per second. Tape speed is 2.8 m/s.

Order of Retrieval of Patents with the Help of a Computer

The order of retrieval of patents can be comprehended best of all with the following example. The specialist (designer or developer) asks for a report on what numbers of patent descriptions are in France on the technical question in which he is interested

The operator of the section in the beginning determines to which tabulyagram and to which of its zones the given technical question belongs. From the tabulyagram number it finds the corresponding punched tape. After that machine memory is fed data of patents of the whole zone to which this question belongs. A data-processing program is also introduced. Control is transferred to the beginning of the program.

The program examines all numbers of countries. And if the

France number is found, then the number of the patent next to the number of this country will be printed on paper. Thus there is examined all information of the given zone and all patents belonging to the French fund on the technical question asked will be printed on paper. In such a sequence there are also processed all other inquiries of specialists.

Obtaining, thus, the reference with the number of the patents, the specialist analyzes and studies these patents.

The Technical Effect Obtained from the Application of Machine Retrieval

1. The main effect of the application of machine retrieval of patent descriptions is considerable acceleration of retrieval of patents. For example, retrieving 100 patent descriptions on any technical question with the help of tabulyagrams and recording only the numbers of patents takes a minimum of 50 min. With the help of the machine this work takes no longer than 4 min (input of program and information 2-3 min and retrieval 1 min).

2. The second effect is the releasing of highly skilled specialists from retrieval of patents. This work on their assignment can now be fulfilled by an operator or laboratory technician on a computer.

3. Mechanization of retrieval makes it possible to organize patent-licensing work in reference to the flow chart of development of new articles — in all its stages.

Region of Application

Introduction of the machine method of retrieval of patent descriptions is easy to carry out in all organizations having ETsVM. One practical application of the machine method of retrieval of patents is expedient in those cases in which the frequency of retrievals is high — 10-15 retrievals per day.

Certain difficulties arise in organizations not having computers. Since in these cases not the problem of use of the machine for retrieval of patents but the problem of its obtaining or the problem of cooperation with other organizations having such machines is solved.

Application of ETsVM for bigger – republic patent funds – is unconditionally expedient, but in these cases problems of systematization of the whole fund come up. All patents have to spread according to narrowly specialized or other criteria.

This is a question of the competence of the central scientific research institutes.

However, these problems must be solved since the study of technical levels according to patent-technical descriptions and the carrying out of patent examinations takes up too much of the time of highly skilled specialists. Reduction of this time is problem number one, all the more so because the volume of information is continuously increasing.

The contemporary level of development of computer technology permits solving this problem already now and with great effect, which is confirmed by the experiment of our organization.

CERTAIN QUESTIONS OF MACHINE PREPARATION OF INDEXES FOR CURRENT AND RETROSPECTIVE RETRIEVAL OF SCIENTIFIC AND TECHNICAL INFORMATION MATERIALS

D. I. Mekhtiyev and E. K. Kuznetsova

In contemporary practice of information service a more and more considerable place is being occupied by various kinds of indexes prepared with the help of punch-card machines and computers, for example, permutational indexes and varieties of them, indexes of bibliographic references, and indexes of the tabular type.

The possibility of mechanization of almost all stages of the technological process of the manufacture of machine indexes ensures the operativeness of their preparation, which is especially important if indexes are used as signal information. The reduction of expenditures of manual labor essentially decreases the periods of preparation of such indexes.

Preparation of bibliographic indexes for current and retrospective retrieval of scientific and technical literature in the practice of libraries and organs of scientific and technical information is labor-consuming work requiring large expenditures of time. Automation of preparation of indexes permits getting current information to the consumer and preparing at the order of specialists in the shortest periods retrospective bibliographic indexes with respect to various branches of science and technology.

This work examines certain questions of the method of creation of a permutational index, embracing 2500 sources in the field of development and exploitation of naval petroleum and gas deposits from 1954 to 1966. The index was prepared by the Azerbaydzhan Institute of Scientific and Technical Information of the Gosplan of the Azerbaydzhan Soviet Socialist Republic in conjunction with the All-Union Institute of Scientific and Technical Information of the State Committee on Science and Technology and the Academy of Sciences of the USSR (Division of Semiotics, leader of work G. E. Vleduts).

During preparation of the index there was used the "Ural-4" and the punch-card equipment in the system.

This index includes various literary sources: journals, collections of scientific and technical information, bulletins, books, articles, and other materials, which illustrate questions of drilling and exploitation of naval oil wells, hydrotechnical construction, corrosion of equipment of naval petroleum industries, economics, and others. It consists of three parts:

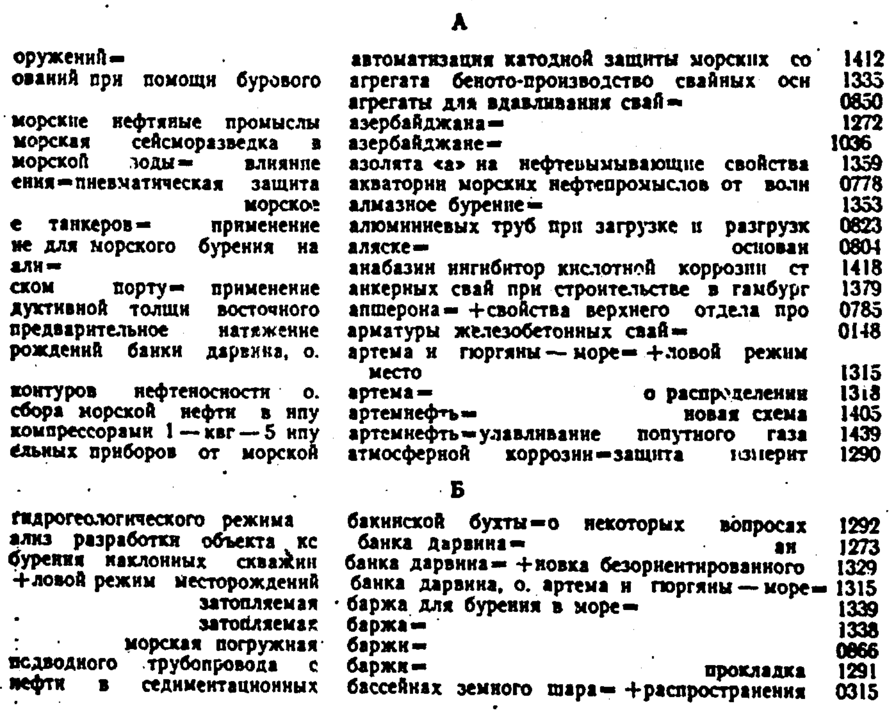
1) the index of key words UKS located in alphabetical order in the input column in approximately the center of the table, and information codes (right column) which are the input into the other two parts;

2) the bibliographical part;

3) the alphabetical index;

On the left and on the right of the key word in the line there is placed the context of the title, definitizing its meaning (see table).

Sample Index of Key Words [in Russian alphabetical order. Right context continues at left margin].



А

оружений=	автоматизация катодной защиты морских со	1412
ований при помощи бурового	агрегата беното-производство свайных осн	1335
	агрегаты для вдавливания свай=	0850
морские нефтяные промыслы	азербайджана=	1272
морская сейсморазведка в	азербайджане=	1036
морской зоды= влияние	азолята «а» на нефтевымывающие свойства	1359
ения=пневматическая защита	акватории морских нефтепромыслов от волн	0778
морское	алмазное бурение=	1353
е танкеров= применение	алюминиевых труб при загрузке и разгрузк	0823
не для морского бурения на	аляске= основан	0804
али=	анабазин ингибитор кислотной коррозии ст	1418
ском порту= применение	анкерных свай при строительстве в гамбург	1379
дуктивной толщи восточного	апшерона= +свойства верхнего отдела про	0785
предварительное натяжение	арматуры железобетонных свай=	0148
рождений банки дарвина, о.	артема и гюргяны — море= +ловой режим место	1315
контуров нефтеносности о.	артема= о распределении	1318
сбора морской нефти в нпу	артемнефть= новая схема	1405
компрессорами 1 — кгг — 5 нпу	артемнефть=улавливание попутного газа	1439
ельных приборов от морской	атмосферной коррозии=защита измерит	1290

Б

гидрогеологического режима	бакинской бухты=о некоторых вопросах	1292
ализ разработки объекта кс	банка дарвина= ан	1273
бурения наклонных скважин	банка дарвина= +новка безориентированного	1329
+ловой режим месторождений	банка дарвина, о. артема и гюргяны — море=	1315
затопляемая	баржа для бурения в море=	1339
затопляемая	баржа=	1338
морская погружная	баржи=	0866
подводного трубопровода с	баржи= прокладка	1291
нефти в седиментационных	бассейнах земного шара= +распространения	0315

During preparation of the index there were carried cut the following operations: separating of key words, composition of a list of nonkey words, and editing of titles of materials. In the list of nonkey words ("empty words") we include conjunctions, prepositions, certain forms of auxiliary verbs, certain adjectives and nouns not carrying basic semantic load in the text of titles, and also the most frequently repeated terms:

Exemplary list of words excluded from index.

- Analysis
- More
- Struggle
- Future
- Faster
- Probable
- Interaction
- Influence
- Quantitative
- Short
- Big
- People
- Measures
- Method
- Powerful
- Certain
- Rules
- During
- Application
- Example
- Principles
- Causes
- Problem
- Production

Exemplary list (Continued).

To the question
For the first time
Gigantic
Data
Action
Permissible
Problems
Regularity
View (from the point)
Change
Study
Investigation
Use
Results

New
Guarantee
Regions
General
Definition
Organization
Peculiarities
First
Perspective
Areas
Preparation
Indices
Position
Therefore

Process
Properties
Conjunctions
State
Method
Periods
Point
Improvement
Conditions
Section
Factors
Characteristic
Integers (in)
Stage

Edited titles of publications were recorded on punched cards according to a preliminarily developed mock-up.

Separating of key words is the most responsible part of the work during preparation of a permutation index and usually should be fulfilled by highly skilled specialists. The quality of this work to a large extent determines the quality of the index. Furthermore, if the index is reloaded with words not bearing the basic semantic load, then retrieval according to UKS will be hampered, since the latter will strongly swell.

It is necessary to note that optimum recording density of a UKS page plays a large role in evaluation of preparation of permutation indexes. The question of volume of index is especially sharp during processing of large collections of publications. In this case speed of preparation of permutation indexes in combination with the economic form of recording is an advantage of such indexes.

Complexity of apportionment of key words is caused by the fact that the titles of publications for the most part do not correspond to the necessary requirements, i.e., as a rule, are multiword and uninformative. As an example let us give the following title:

— "Research in the problem of origin of oil. Hydrocarbons in contemporary deposits." After editing the title took such a form: "Origin of oil. Hydrocarbons in contemporary deposits."

— "Application of a conveyer on a barge for the pouring of dam." After editing — "Conveyer on barge for pouring of dam."

In some cases the title absolutely does not disclose the contents of publications and must be completely replaced. Into the UKS there were also introduced new key words taken from text.

The process of apportionment of key words is considerably simplified in the presence of a ready list of nonkey words and a thesaurus. During preparation of permutation indexes with the use of thesauri and microthesauri on special individual questions of science and technology there is excluded human participation in the marking of titles, and, consequently, there is removed subjectivism in the selection of key words, inasmuch as in a thesaurus there are enveloped all basic descriptors with respect to a branch, taking into account basic genus-species and associative relationships, and others. Standardization of terminology will also allow eliminating one of the most important deficiencies of the permutation index — scattering of information, for example as a result of cross-references.

The first variant of index obtained on the tabulyagram was thoroughly edited. In connection with the fact that the list "of empty words" was composed with certain assumptions, in index there were often encountered contexts considerably truncated on the left and on the right, which hampered perception of the title. In such cases there were omitted terms which in the first editing were left as key terms without the proper base. It was necessary to exclude from the UKS the title of source which started from the dropped key word, and conversely to include in the UKS the title, starting with the new key word; all the words were introduced in alphabetical order.

The average title of a book, journal, article, etc., in the index contains 4 key words and only in rare cases more than that. If the title of a publication has 3-4 key words, then there is room for it in the index for the most part wholly without truncation.

The end of the title, if it fits completely on one line, is marked (:=); however, if the title does not go on one line the length of which is limited (in this case not more than 67 characters without code), then the context is partially broken. The break of context is marked (+).[1] The remaining part of it is considered accepted and is kept in the index if the number of letters of the final word is not less than 4.

If it is impossible fully to decode the title because of the truncation of the context one should find the second key word and if necessary subsequent key words of this title. During the reading of the title (in each individual case with the new key word) it is possible to restore the context truncated in one of the variants since during different key words there are broken various words entering context.

The index starts from the list of words excluded from the titles of sources (so-called "empty words").

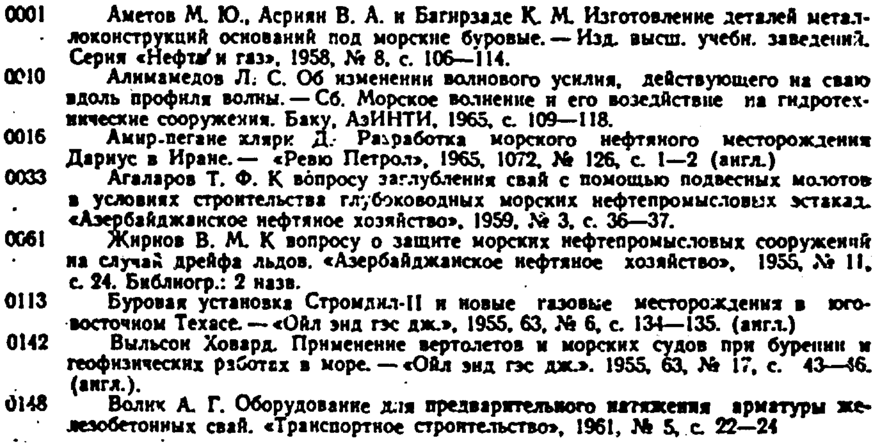
During more thorough examination and editing of titles it would have been possible to separate an additional series of nonkey words. However, considering the well-known labor consumption of such editing, and also the fact that presence in the index of an insignificant number of these words will not elicit special inconveniences during the use of the index, certain assumptions are made.

After the list of words excluded from the index there is placed

[1] There are given signs used in test output of the permutation index on development and exploitation of naval petroleum and gas deposits (360 publications), collected by the typographical method. Full output of the index has other signs.

the "Index of Key Words." Then there follows a bibliographical list of sources, in which the authors and titles of all sources and others are shown.

Sample Bibliographical Index.



0001 Аметов М. Ю., Асриян В. А. и Багирзаде К. М. Изготовление деталей металлоконструкций оснований под морские буровые. — Изд. высш. учебн. заведений. Серия «Нефть и газ», 1958, № 8, с. 106—114.

0010 Алимамедов Л. С. Об изменении волнового усилия, действующего на сваю вдоль профиля волны. — Сб. Морское волнение и его воздействие на гидротехнические сооружения. Баку, АзИНТИ, 1965, с. 109—118.

0016 Амир-пегане хлярк Д. Разработка морского нефтяного месторождения Дариус в Иране. — «Ревю Петрол», 1965, 1072, № 126, с. 1—2 (англ.)

0033 Агаларов Т. Ф. К вопросу заглубления свай с помощью подвесных молотов в условиях строительства глубоководных морских нефтепромысловых эстакад. «Азербайджанское нефтяное хозяйство», 1959, № 3, с. 36—37.

0061 Жирнов В. М. К вопросу о защите морских нефтепромысловых сооружений на случай дрейфа льдов. «Азербайджанское нефтяное хозяйство», 1955, № 11, с. 24. Библиогр.: 2 назв.

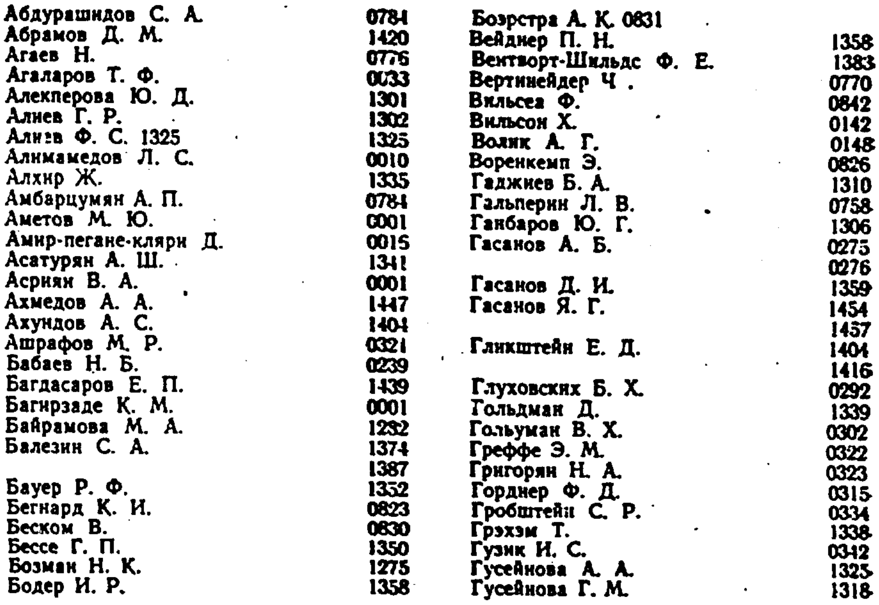
0113 Буровая установка Стромдил-II и новые газовые месторождения в юго-восточном Техасе. — «Ойл энд гэс дж.», 1955, 63, № 6, с. 134—135. (англ.)

0142 Вильсон Ховард. Применение вертолетов и морских судов при бурении и геофизических работах в море. — «Ойл энд гэс дж.». 1955, 63, № 17, с. 43—46. (англ.).

0148 Волик А. Г. Оборудование для предварительного натяжения арматуры железобетонных свай. «Транспортное строительство», 1961, № 5, с. 22—24

On the left tnere is a column of numbers indicating the code for every work. The index is completed by an alphabetical list of authors.

Sample Author Index [in Russian alphabetical order].



Абдурашидов С. А.	0784
Абрамов Д. М.	1420
Агаев Н.	0776
Агаларов Т. Ф.	0033
Алекперова Ю. Д.	1301
Алиев Г. Р.	1302
Алиев Ф. С. 1325	1325
Алимамедов Л. С.	0010
Алхир Ж.	1335
Амбарцумян А. П.	0784
Аметов М. Ю.	0001
Амир-пегане-кляри Д.	0015
Асатурян А. Ш.	1341
Асриян В. А.	0001
Ахмедов А. А.	1447
Ахундов А. С.	1404
Ашрафов М. Р.	0321
Бабаев Н. Б.	0239
Багдасаров Е. П.	1439
Багирзаде К. М.	0001
Байрамова М. А.	1232
Балезин С. А.	1374 1387
Бауер Р. Ф.	1352
Бегнард К. И.	0823
Беском В.	0830
Бессе Г. П.	1350
Бозман Н. К.	1275
Бодер И. Р.	1358
Боэрстра А. К. 0831	
Вейднер П. Н.	1358
Вентворт-Шильдс Ф. Е.	1383
Вертинейдер Ч .	0770
Вильсеа Ф.	0842
Вильсон Х.	0142
Волик А. Г.	0148
Воренкемп Э.	0826
Гаджиев Б. А.	1310
Гальперин Л. В.	0758
Ганбаров Ю. Г.	1306
Гасанов А. Б.	0275 0276
Гасанов Д. И.	1359
Гасанов Я. Г.	1454 1457
Гликштейн Е. Д.	1404 1416
Глуховских Б. Х.	0292
Гольдман Д.	1339
Гольуман В. Х.	0302
Греффе Э. М.	0322
Григорян Н. А.	0323
Горднер Ф. Д.	0315
Гробштейн С. Р.	0334
Грэхэм Т.	1338
Гузик И. С.	0342
Гусейнова А. А.	1325
Гусейнова Г. М.	1318

Retrieval is carried out in the following way. In the division "Index of Key Words" along the vertical there is examined the alphabetic list of key words and on the right along the horizontal there is noted the information code of the source of interest to the specialist. From the information code, which is the input of the bibliographical part of the index there is found the name of the source, the year of publication, etc. The author's index serves also for transition to the bibliographical part of the permutation index. For example, there is given the title coded 1404 of one and the same source, consisting of four key words: "Electrochemical protection of hydrotechnical constructions from corrosion," and diagram of transition from index of key words to bibliographical and author's index (Fig. 1, see above).

So that it is more convenient to use the index and in order to expand the aspects of retrieval the permutation index could additionally have alphabetically coded names of sources, year of publication, and others. In this case as a large number of the most diverse forms of information sources were used, this possibility is not used. Compilers of the index pursued another goal – on a diagram of the permutation index, simple in structure, to reveal its essence.

In the index there, were used four-digit codes which in the process of machine processing in all cases are placed in the UKS in the column on the right. So that it is more convenient to pass from the index of key words to the bibliographical part of the index in this part the code is given in the column on the left. In the author's index the code is used not for retrieval but for transition to the bibliographical part; therefore, it is located in the column on the right.

Recordings were processed on the "Ural-4" according to a specially composed program: reproduction of titles according to the number of noted key words, intramachine sorting of reproduced titles according to the alphabet of key words in the input column of the UKS (28th) and printing.

For the given permutation index the author's and bibliographical parts were prepared manually. Their processing on ETsVM does not present any principle difficulties. In this case all parts — UKS and bibliographical and author's index — after machine processing are printed separately; at the output of the ETsVM we have three tabulyagrams which can be used directly for subsequent reproduction by the photo-offset method.

In connection with the fact that during retrieval of sources by this index there is used visual examination of titles, the alphabetical list of authors or the bibliography, much attention should be paid to shaping and printing the index. During preparation of the test output of the permutation index according to the development and exploitation of naval petroleum and gas deposits there was proposed using the tabulyagram for direct photographing without reprinting on the machine and subsequent reproduction by the photo-offset method. However, in connection with the fact that there was required additional editing of the tabulyagram, and the quality of "Ural-4" printing was low, it was necessary to issue the index by the usual typographic method (typesetting on linotype and printing on a flat-bed machine). In order to facilitate visual search the part of the context to the left of the list of key words and the code part of the index are printed on a colored background.

When the permutation index is prepared as signal information, the tabulyagram should nevertheless be used directly for printing. During composition of the index on literary sources after a considerable interval of time for retrospective retrieval there is justified use of traditional forms of printing (in particular, flat printing). It is true that periods of output in this case are somewhat increased; however, the permutation index issued by such a method is very convenient to use. Considering that one of the main advantages of the permutation index are compressed periods of its output, subsequently efforts should be directed towards improvement of the quality of printing at the output of ETsVM, and introduction of all possible devices and attachments which allow improving the quality of printing of tabulyagrams and accordingly turning to publication of permutation indexes and for retrospective retrieval directly from the tabulyagram by the offset method.

In contrast to foreign permutation indexes on every source there is proposed giving an index of universal decimal classification. This will allow partially removing the basic deficiency of the index, connected with incomplete opening of the contents of the source according to the title, and will also accelerate the process of subsequent delivery of the source. Furthermore, considering that the most widespread form of indexing of literary sources is Universal Decimal Classification, the permutation index prepared with indices of this classification will be easy to grasp by specialists and library workers and will be widely used.

In this case there was no tendency to use Universal Decimal Classification as a exploration system. However, in the permutation index there is a real possibility of using Universal Decimal Classification for these purposes. The method of retrieval according to Universal Decimal Classification and questions of location of indices and their encoding have to be solved in the process of preparation of the index, taking into account creation of maximum conveniences for retrieval.

The main advantage of permutation indexes is the great gain in time during preparation of indexes with the help of ETsVM. This is especially noticeable during comparison of time needed for their preparation and with time of composition of subject indexes in library practice.

The basic time during composition of permutation index is spent preparing data for input by external devices. In particular, time is mostly expended punching and subsequently checking holes on punch-card equipment. From available data during mechanized processing of material, 70% of the time is spent punching and checking. At the same time during application of ETsVM labor of punching and subsequently checking is 90% of the total labor needed for machine data processing. It is sufficient to say that it takes about 17 minutes to prepare a permutation index of 300-400 titles in the presence of 600 nonkey words on the IBM 7090.

Creation and wide use of permutation indexes will allow operationally in more compressed periods informing consumers of information about the latest publications, increasing the effectiveness of current bibliography and making it accessible to all categories of specialists.

On the example of a permutation index in the region of development and exploitation of naval petroleum and gas deposits it is possible also to conclude that analogous indexes for retrospective retrieval of literary sources is highly effective. Creation of such indexes on literary sources on different branches of science and technology for defined periods of time will render priceless help to specialists.

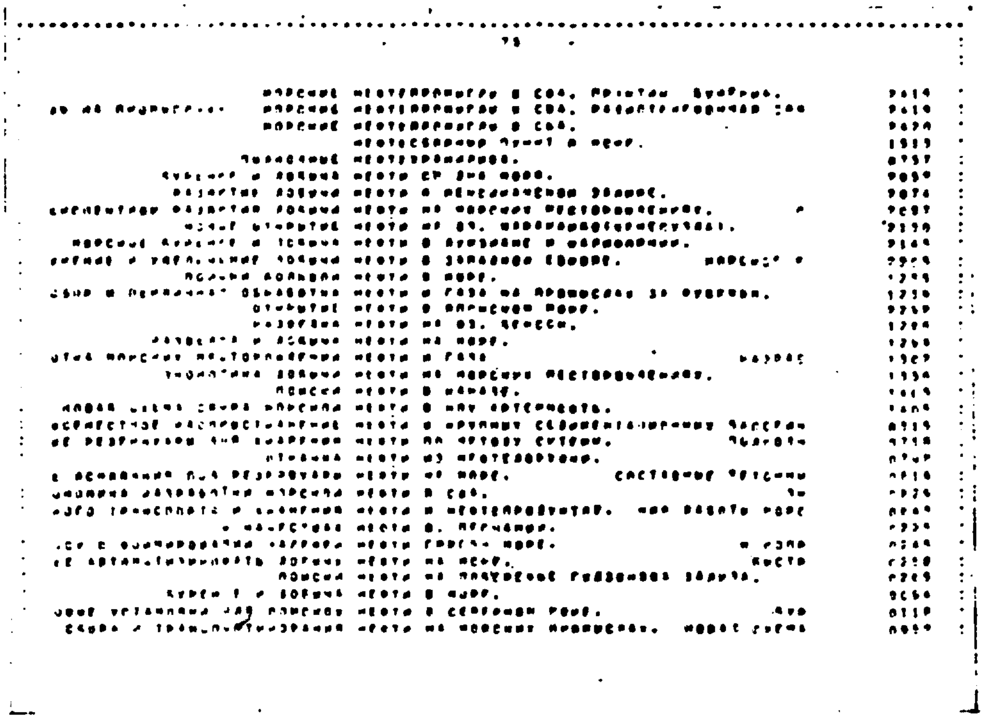


Fig. 2. Tabulyagram UKS.

At present we are conducting experiments for the purpose of further improving permutation indexes.

There is prepared an index of the type "KWOC" with the use of the "Minsk-22." Experience accumulated in the course of the developments described permits formulating the question of creation of fundamentally new indexes based on the indexes examined above, absent in foreign and domestic practice.

Bibliography

1. Mikhaylov A. I., Chernyy A. I., Gilyarevskiy R. S. Osnovy nauchnoy informatsii (Bases of scientific information). Izd-vo "Nauka", 1965, 654 s.

2. Vleduts G. E., Stokolova N. A. O sostavlenii i ispol'zovanii ukazateley tipa "klyuchevykh slov v kontekste" (On composition and use of indexes of the "key words in context" type). "NTI", 1964, No. 4, 30-35.

3. Kuznetsova E. K. Permutatsionnyye ukazateli (Permutation indexes). "NTI", 1966, No. 6, 28-37.

4. Voress H. E., Improvements in a permuted title index. "Amer. Docum.", 1965, 16, No. 2, 97-100.

5. Haas A. K. Internal alerting with keyword in-context indexes. "J. Chem. Docum.", 1965, 5, No. 3, 160-163.

MECHANIZED IPS FOR A VARIED FUND DEVELOPED AND INTRODUCED BY THE UPPER-VOLGA TsBTI

V. M. Mesnik

At the Upper-Volga TsBTI [expansion of acronym unknown] in accordance with the plan of development of the national economy of the RSFSR from 1964 to 1965 there was conducted scientific research on mechanization of processing, retrieval, and reproduction of scientific and technical information, which in 1965 was introduced into the work of the TsBTI.

Naturally, the ideal information service cannot be immediately created. Up to now no such service exists anywhere in the world. However, in any case with application of mechanization urgent questions of organization of information service today can be solved more expediently, profitably, and quickly than without it.

Under conditions of a regionally varied organ of information, only domestic punch-card equipment was recognized by us as accessible and expedient for realizing the problem at hand (Penzenskiy plant TEM [expansion of acronym ambiguous]).

Reproduction and removal copies is carried out on domestic copiers and microfilming equipment. There are used:

- Punch P80-6

- Verifier K80-6;

- Sorter S80-5M;

- Installation for microfilming UDM-2;

- Electrographic reproduction apparatus ERA-2F;

- Thermocopier;

- Microphotographer.

As information carrier in the system of mechanized retrieval there is used an 80-column punched card "Glavmekhscheta."

As a basis of the retrieval system for mechanized retrieval under conditions of a varied fund we assume a language already created — universal, international, and with relatively well-developed logical links between ideas, i.e., a language of Universal Decimal Classification.

And only those ideas which are not reflected in tables of Universal Decimal Classification are coded with the help of digital code of another type, which is created in the division. Brands of machines and apparatuses, their parameters, and certain other characteristics will be thus encoded. Conditionally this code is called "descriptor," although traditional use of this term has another meaning.

It is known that Universal Decimal Classification is assumed as a basis of retrieval language in systems of mechanized retrieval successfully used in Czechoslovakia, Hungary, and other countries.

However, use of Universal Decimal Classification in a system of mechanized retrieval involves a number of difficulties: 1) the same questions have different indices if they belong to different regions of application; 2) coding of complicated questions allows arbitrary arrangement of indices corresponding to subjective thinking of indexers.

All of this in a system of mechanized retrieval must lead to serious information loss and a high percentage of information "noise." Therefore, using Universal Decimal Classification – the traditional system of classification for mechanized retrieval – we tried to establish strict methodical rules obligatory during indexing of materials with respect to four-five aspects. By these rules every retrieval criterion of a document is coded always in a definite zone of a punched card. Thus there is established a single diagram of construction of a complicated and compound index for all coders, and during composition of the retrieval program it is always absolutely obvious in what zone a given retrieval criterion is coded (See Appendix).

A code for mechanized retrieval includes:

1. An index of Universal Decimal Classification, complete and exact, with use of all possibilities of the system.

2. A "descriptor" which expresses brands of equipment and their parameters, i.e., those characteristics not in Universal Decimal Classification.

3. Year of publication of information.

Code for mechanized retrieval takes up 33 columns on an 80-column card. The remaining part, in which it is possible to dispose 1000 printed characters, is used for recording the text of the abstract. The field for punching code is divided into five zones, in each of which there is always recorded a strictly defined part of the code.

1) I zone – 12 columns. There is coded technological process and equipment – all these ideas are expressed by the basic index of Universal Decimal Classification and special determinants of type 0;

2) II zone — 6 columns (13-18): definitizing characteristics (questions of checking, regulation, and machine parts) are expressed through determinants of type 00 and of type - (hyphen);

3) III zone — 9 columns (19,227): object of processing, region of application, material — all of this is expressed by the index of Universal Decimal Classification after the ratio sign (:);

4) IV zone — 5 columns (28-32) brands of machines and apparatuses, parameters, and other characteristics not in Universal Decimal Classification. (See mock-up of punched card in Fig. 1.)

I zone — 12 columns	II zone — 6 columns	III zone — 9 columns	IV zone — 5 columns	V-1	VI zone
Basic index of Universal Decimal Classification and special determinants of type .0	Determinants of type .00 and -	Index of Universal decimal classification after sign of ratio	Descriptor	Year of publication	Text
621 388 16 032^1	-0.3.35^1	615^6		5	

Fig. 1.

Such division of code into zones in which each part of it is always punched in certain columns of the punched card, creates important advantages during multiaspect retrieval, allowing inquiries both general and special in nature to be answered.

If the inquiry is received to issue materials on welding apparatuses, it is possible to examine the whole array on welding only with respect to the 1st zone 621.791.75.03.

If the user is interested in information on automatic welding apparatuses, the 1st and 2nd zones of the same array.

In order to obtain information on the make of equipment it is sufficient to examine only the descriptor zone of the corresponding array.

Let us examine the method of composition of the descriptor dictionary:

Classifying information is material in which the make of equipment is shown, the coder records it in the descriptor dictionary. In this dictionary every letter is designated by corresponding figure A — 01, B — 02...K — 10, P — 15, etc. The coded idea is recorded by initial letter or figure "Welding automation" ASID-3M on "A," indicator instrument PKSh5 on P.

Digital expression of the descriptor consists of the reference number of its initial letter and the reference number of the given recording on this letter. Thus, ASID-3M (make of welding automatic machine) was coded 0116, where 01 is the designation of letter "a," and 16 is the reference number of recording. AT-4-120-1 is coded 01.29 according to the same principle.

When the corresponding recording is made in the descriptor dictionary and the make of equipment is digitally coded, there is ordered a card on which there is designated the alphabetical or numerical expression of the make, the name of the equipment, the code appropriated to it, and the index of Universal Decimal Classification. This card is placed in the card file of the descriptors in strictly alphabetical order by name of make, without taking the digital expression of the code into consideration.

During composition of the retrieval program there is used already only this card file (and not a dictionary), since from it it is easy to find the make of any equipment information about which is in the fund and the code appropriated to it and from the code during several minutes of sorting the information is detected.

The index of Universal Decimal Classification entered on the card indicates what division should be examined on the sorter.

Retrieval time on the sorter is 2-3 minutes. Today this fully organizes us, the more so because there is no longer a need for manual sorting and arrangement of maps — labor-consuming and thankless work.

Another method of preventing subjective approach of the indexer to selection of key words and a method of their expression by Universal Decimal Classification is a card file of decisions in which there are fixed all decisions made while documents are being processed. A sample card from the card file of decisions is shown below.

Casting under pressure 621.74.043

— created compressed air or gas 621.74.043.3

— equipment 621.74.043.3.06

— pressing pneumatic cylinder 621.74.043.3.06-222

— created mechanically 621.74.043.2

— assembly line 621.74.043.2:658.527

As the basis of this alphabetical-subject card file there is assumed an index. However, even in this question there is deviation from traditional library procedures: in the card file there is recorded the key word encountered during indexing of the document — the procedure utilized in systems of the descriptor type — but in contrast to these systems to the given word there is appropriated ready international digital code — the index of Universal Decimal Classification having its own definite place in the general hierarchical system of classification of technical ideas — and, therefore, is connected and with a more general technical idea and with its narrow questions.

Welding apparatus — 621.791.75.03, where 621.791 is welding, and 621.791.75 is electric arc welding.

On the other hand, with such a method of management the card file of decisions acquires an orienting function: presence in card file of decisions of the card with the key word is evidence that the corresponding document is in the fund.

Management of this card file is labor-consuming work, but it justifies itself, all the more so since rates of growth of the fund and this card file are incommensurable, to which Fig. 2 testifies.

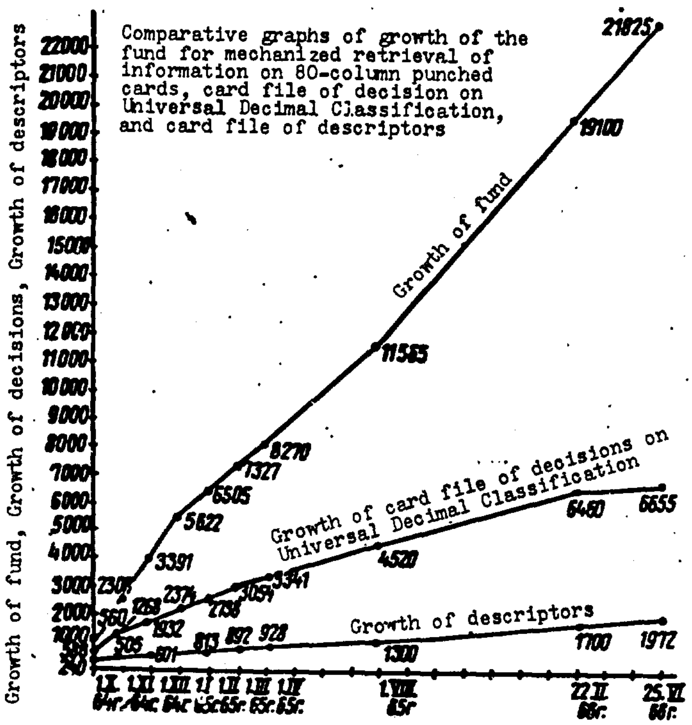

Comparative graphs of growth of the fund for mechanized retrieval of information on 80-column punched cards, card file of decision on Universal Decimal Classification, and card file of descriptors

Fig. 2.

When in the card file there is reflected all or almost all the terminology of the given branch, the process of indexation will be essentially facilitated.

We experimentally checked a system having ah array of 10 thousand punched cards. Very reassuring results are obtained.

A table of these results is given below.

Subject of inquiry	Array in which retrieval was carried out	Number of cards of array answering inquiry, a	Amount of issued information answering inquiry, B	Quantity of unissued information C	Quantity of issued information answering inquiries	% noise $\frac{a\cdot 100}{B}$	% unissued information $\frac{C\cdot 100}{a}$	Coefficient of completeness of retrieval $\frac{B}{a}$	Coefficient of completeness of retrieval $\frac{B+C}{a}$
Presses for wood processing	1820	13	13	—	—	0	0	1,00	1,00
Prevmotransport*	1450	14	11	3	—	0	21.5	0,79	0,79
Autostripper of yarn for spinning machines	423	5	5	—	—	0	0	1,00	1,00
Lowering of waste on «uslovyaral'noy» machine	371	1	1	—	—	0	0	1,00	1,00
Automatic lines in machine building	4530	61	60	1	—	0	1,6	0,98	0,98
Drying of wood	420	20	15	5	—	0	25,0	0,75	0,75
Zincing	996	28	21	7	—	0	25	0,75	0,75
Application of wood plastics of pressed wood	5170	6	6	—	—	0	0	100	100
Altogether:		148	132	16	—	0	10,8	0,89	0,89

*Translator's Note: No translation found.

Zero % noise — in other words, among information materials issued by sorting there was not one superfluous document not answering the given question.

The percentage of unissued information from what is available in the fund on a given question of information was 10.8. In the theory and practice of scientific and technical information this index is considered very good.

Of the 16 punched cards unissued 12 were not issued because of the imperfect construction of the sorting. Sampling on a many-valued criterion is carried out by commutation of sockets of the first and second row of the panel "of set" with the help of switching cords, which not always ensures reliable contact and does not give required speed. And from this there is increased the percentage of noise and the percentage of unissued information.

For sampling of information on a many-valued criterion a push-button panel of set is necessary. The deficiency of sorting is the fact that it samples on 12 columns in one run. And for retrieval of information there is necessary an electronic sorter to carry out sampling immediately on all columns.

Usually the following arguments are made against use of Universal Decimal Classification for mechanized retrieval:

a) technical difficulties of indexation;

b) high cost of input;

c) impossibility of expressing the contents of the document in detail;

d) subjectivism of indexer.

How we minimize the subjectivism of the indexer, using Universal Decimal Classification, has already been discussed above.

Indexation by Universal Decimal Classification is a process requiring tense attention and definite knowledge, like any intellectual labor in general. And all engineers of average skill master it in a comparatively short period and index materials with sufficient completeness and accuracy and comparatively rapidly, averaging 25 documents in a day.

In this connection it is appropriate to talk about cost of input, more exactly, the cost of coding one document. It costs us 13-16 kopecks. All remaining consumptions on input are not related to the retrieval language.

If one considers that regional reference and information funds are in considerable measure completed with cards of central branch institutes, which proceed to us with indices of Universal Decimal Classification, then for laying of them in the fund for mechanized retrieval it is necessary only to convert the index into code, and this takes minutes, of course, if in central branch institutes it pertains to indexation with sufficient conscientiousness. Now, unfortunately, this is not observed.

Regarding the degree of fullness and accuracy of coding, although in principle systems of the descriptor type in this respect have to have evident advantages over a system of the classification type, (Universal Decimal Classification) in those systems with which it was possible for us to become acquainted, there is not yet any such advantage: in them lexical composition is too poor, and logical couplings are almost not developed. Moreover all of them are created for service of the fund on a more or less narrow question. For a regional fund a universal system is needed, among which Universal Decimal Classification is undoubtedly best.

What are the methods of processing, storage, retrieval, and delivery of information used in SIF TsBTI?

In the fund of SIF there is put information both proceeding "from the bottom upwards" and obtained "from the top downwards."

According to materials of information cards of enterprises of three regions workers of branch divisions and SIF TsBTI monthly compose lists of proceeding information cards. They are printed on rotoprint in the form of pamphlets with the possibility of cutting them into separate cards for convenient laying in the fund.

Information is printed on the "Optima" (IGV-2) typewriter and carries the following information:

a) name of innovation,

b) by whom developed,

c) date and place of introduction,

d) source of information,

e) annotation,

f) economic effect.

After reproduction of the bits on printer-copier equipment and distribution of them to the enterprises of the economic region, the first copy of it is cut and transferred on the ERA-2F to punched cards.

In a shift one ERA-2F transfers information to 300 punched cards. On information materials entering the TsBTI from other TsBTI and central branch institutes, there is placed in the fund a punched card with a bibliographic description.

The cost of one punched card with transferred text of information is 1.8 kopecks.

Punched cards with transferred information are punched. Punched cards in the receiving pocket of the P80-6 punch are filled, and the operator punches the assigned code.

The productivity of the punch is 1600 punched cards per shift. The punched cards are checked.

They are checked on the K80-6 verifier on the keyboard of which there is a second time collected code according to a mock-up of punching.

Correctness of punching governs the result of subsequent retrieval of inquired information.

The productivity of the verifier is 1600 punched cards per shift. We do not use it completely. Information in the fund is retrieved in such order: in accordance with the subject of the inquiry there is composed a program for mechanized retrieval with respect to the card file of decisions and the card file of descriptors.

The sorter examines the corresponding array of punched cards, documents issued by it are analyzed by an engineer, and then copier-reproducer or microfilming equipment makes copies in 2-3 min, which are sent to the user.

Operational Communication

For operational usé of the latest achievements of science and technology there are necessary not only mechanization of preparation, retrieval, and delivery of information, but also its fast transmission to users.

According to measures provided for by the resolution of the Council of Ministers of the USSR from 1 August 1963 No. 843, organs of information are obliged to organize the obtaining and operational transmission of information with the use of the latest instruments and equipment of communication.

In the Upper-Volga TsBTI since March 1964 there has been teletype communication between division SIF and DNTIP in the cities of Yaroslavl, Vladimir, and Kostroma. There has been established communication with many enterprises of economic regions.

At present there has been established teletype communication with many republic institutes and TsBTI of many economic regions of the country. The teletype communication permits considerably shortening the time it takes to obtain and deliver information.

If the time it takes to obtain information (inquiry – answer) by letter is 10-15 days at best, then with the help of teletype communication it is shortened to 2-3 h (and in certain cases to 20-30 min).

In 1965 310 references were obtained and issued by teletype. Subsequently there is contemplated still wider use of teletype communication for the needs of information.

Conclusion

I. The IPS developed by us is applied in practical activity of varied SIF during two years.

Mechanized retrieval and delivery of information with the use of comparatively cheap punch-card equipment considerably reduces time necessary for retrieval of documents and facilitates the labor of workers of the section.

And although retrieval is carried out according to a limited number of retrieval criteria (4-5) practically in the overwhelming majority of cases we are satisfied with its results.

APPENDIX

TEMPORARY INSTRUCTION ON SEPARATE QUESTIONS OF INDEXATION BY UNIVERSAL DECIMAL CLASSIFICATION AND CODING OF INFORMATION MATERIALS FOR MECHANIZED RETRIEVAL

During coding of information materials one should observe the following rules:

1. The basic index of Universal Decimal Classification disposed in the 1st zone should express the technological process or equipment of which it is a question in the information, and the part of the index which comes after the sign of ratio (i.e., in the 3rd zone) determines the object of processing.

For example: 1. Dyeing of tissues made of synthetic fibers 677.842:677.494.064.

2. Milling machine for worm cutting of straight-tooth bevel gears 621.914.5:621.833.22.

Only in those cases in which it is impossible to express the technological process by the basic index of Universal Decimal Classification is the object of production placed in the first zone, and the process and equipment designated by determinant .002...

For example: 1. Zapletka [no translation found] of cables in production of truck-cranes 621.86.065.3.002.72.

2. If the contents of the information can be expressed only with the use of two signs of ratio, one should give dipole cards (the first idea with respect to the second and the first idea with respect to the third). For example: "Sharpening" of a drill with an artificial diamond 621.923.6:621.951.4:666.233.

code: 1) 621.923.6^{11}621.951.4......6

2) 621.923.6^{11}666.233......6

3. In a number of cases very important technical ideas on Universal Decimal Classification can be expressed only through a determinant of type .0.

QUESTIONS OF DEVELOPMENT OF AN INFORMATION-RETRIEVAL SYSTEM FOR AUTOMATED CONTROL OF THE WORK OF A NAVAL FLEET OF STEAM NAVIGATION

A. N. Kiselev and I. A. Mikhaylova

Further improvement of the control system of a naval fleet of steam navigation is possible only on the basis of introduction of mathematical methods and means of computer technology and a way of gradual transition to a system of automated control (SAU). Such a system can take over the following basic functions:

a) automated collection, treatment and storage of information about state of transport process;

b) sizing up of the situation put together in the transport process and development of recommendations on its regulation;

c) machine preparation of initial data and solution of concrete problems of control of the work of the fleet;

d) preparation and delivery of answers to inquiries of management apparatus and reports on set forms, etc.

Naturally, the SAU will be created in several stages, distinguished by a circle of solved problems, degree of automation input and data processing, structure of controls, etc.

The material of the present report is part of large overall research in information provision of SAU. This complex also establishes the composition of information participating in solution of the problems of control of the work of the fleet; its standardization and unification, development of extramachine and intramachine languages and forms of presentation of information, organization of transmission of information on different communication channels, development of founded requirements for composition of technical means, and provision of the necessary sequence and continuity of treatment and transmission of information.

The information-retrieval system (IPS) for automated control of the work of the fleet is developed in the following sequence:

1. Analysis of the complex of problems of control, their information communications, and peculiarities of solution.

2. Foundation of SAU in the form of an IPS variant. Peculiarities caused by assignment and conditions of work of the system.

3. Development of a system of inquiries and IPS (composition of current forms, inquiries of management apparatus, preparation of initial data for solution of concrete problems of control).

4. Classification of objects and their characteristics. Determination of composition and volume of arrays of information (information tables).

5. Development of principles of organization of storage and restoration of information.

6. Development of method of distribution of information in memory unit (ZU).

7. Development of algorithms and programs of input and restoration of information.

8. Development of algorithms and programs of preparation of initial data of solution of problems of control of the work of the fleet.

9. Development of algorithms and programs of answers to inquiries.

The economicomathematical model of the system of automated control of the work of a fleet of steam navigation has more than 80 different problems, which can be conditionally divided into three basic groups.

The first group includes annual and quarterly planning of work of fleet, loading of ports and ship repair, composition of the optimum diagram of travel and arrangement of vessels by lines and directions, etc.

The second group consists of problem of operational planning and regulation of the work of the fleet, development of a monthly work schedule of vessels, scheduled assignment to a concrete vessel, appraisal of the daily situation and submitting recommendations on readdressing of vessels, etc.

In the third group there are problems of operational and statistical calculation and analysis of the work of the fleet.

For solution of a large complex of problems on control of the work of the fleet it is necessary to process and store in EVTsM a large volume of information: plans of works, data about the current state of the transport process, position of vessels, state of ports, presence of cargoes, technical characteristics of vessels and ports, tariffs on transport, norm of treatment of vessels in ports, data about completed trips, etc.

Analysis of information participating in the solution of the complex of problems of automation of control of the work of the fleet bears witness to their large information communication. In a majority of problems the same initial intermediate data are used, and the results of one problem are source data for solution of other problems, etc. Thus, for example, in problems of composition of a chart of work of vessels, scheduled assignment, composition of a cargo plan, and readdressing of vessels there participate such data as load capacity, load-carrying capacity, speed of movement of vessel, distance between ports, tariffs on transport of cargoes, etc.

Experience in solving individual problems shows that manual preparation of initial data for solution of problems of control, their retrieval in documents, distribution in a definite sequence, punching and input takes 90-95% of the time spent solving the problem.

Part of the problems intended for daily solution, have limitation with respect to time of solution. The first of these is the problem of regulation of the work of the fleet, which must be solved in a few hours. Furthermore, daily there will be solved the problem of composition of scheduled assignment for vessels finishing a trip, composition of a cargo plan of vessels before loading, delivery of data on operational account, and delivery of reference data on the state of transport process in different aspects. Manual preparation of initial data for solution of these problems requires large expenditures of working time of technical personnel and loads auxiliary equipment.

The biggest effect in this case can be obtained if the system of information is formed taking into account interaction of problems of information flows.

During the analysis of the complex of problems of control, besides the establishing of information links between them and the conditions of their solution, there is determined the composition of the necessary information. For every element of information (index) there are indicated corresponding characteristics (unit, accuracy of measurement, maximum value, periodicity of changes, etc.).

Storage of information and its distribution in the memory units of the computer can be organized in two ways.

In the first case immediately after input special arrays are formed each of which is useful for solution of a certain problem. Creation of such arrays is acceptable if the complex of problems is known in advance and small. If the complex of problems is large, there will be a greater number of arrays, where the same data can appear in a number of arrays, which creates certain difficulties. At present certain problems use small volume of data and are solved for a concrete group of vessels or individual vessels, for example, the problem of composition of scheduled assignment and the composition of the cargo plan of the vessel. Big steamship companies unite hundreds of vessels; therefore, the composition of arrays of information for every problem and in reference to concrete vessels and ports is bulky and irrational.

A second, more rational method is presentation of SAU in the form of an information-retrieval system and organization of information in the form of information tables.

During foundation of SAU in the form of an information-retrieval system there were used theoretical developments of N. A. Krinitskiy, O. A. Abramov, and others [1, 2, 3].

Information systems describe the state of certain objects in their interconnection. An information system is a set of sources and a depository of information, a collection of algorithms of selection of information, and an information-carrying system.

An information control system of a fleet can be considered a particular case of automated systems possessing a number of specific properties governed by the purpose and conditions of its work. The system of information provision of SAU possesses criteria of a dynamic information-retrieval system:

1) restoration of information about the state of objects participating in the transport process and conditions of shipments;

2) distribution of information according to a definite law;

3) preparation of initial data for solution of problems of control;

4) composition of answers to inquiries.

The first basic peculiarity of this IPS is limited composition of information elements in it. The main elements of the transport process are vessels, ports, and loads. Knowing the perspective of development of the fleet, it is possible to count the assigned number of vessels of steam navigation in a certain interval of time, for example, a year. The composition of world ports changes more and more rarely. The product-list of transported cargoes also changes rarely.

A majority of inquiries to the information system are known beforehand and can be rigidly programmed. The obtaining of an answer is the delivery of information according to a standard program. There is no need in this case to translate the inquiry from input language into information language since information at the input to the system will be basically presented in standardized language. Information should be issued in convenient human-readable form, for example, in standardized Russian.

The information-retrieval system will conduct a calculation of the work of the fleet, both operational and statistical, practically without documentation, having freed management from unnecessary tiresome work on data processing and composition of various reports and references.

Therefore, an important question during development of IPS is analysis of the work of management apparatus of steam navigation and problems of control and investigation of inquiries.

To such inquiries there can be referred delivery of information about the position of vessels of steam navigation at a definite moment of time, the presence of vessels in a certain region, daily or monthly transported loads, fulfillment of the scheduled assignment by a vessel, the budget of time of vessels in a calendar cut, etc.

Every object of the control system can be described by a set of characteristics X_1, X_2, ... $X_{\text{ж}}$. Moreover in the information system there are stored concrete values of these characteristics. The characteristics of an object are not only numerical values of any parameter of the object, but also other data determining the participation of this element in the transport process. In this case the concrete value of characteristics can be represented not only by a numerical value, but can also have the form of a certain alphabetical equivalent, word or even word combination. Information about the objects of the system is in the form of information tables. Objects are unified into information tables, taking into account convenience of retrieval of data during solution of a problem.

Information systems are created for production of information about a certain set of objects M. Every object should be assigned a list of properties. A set of M objects is divided into several nonintersecting classes. Every class can also be split into subclasses, sub-subclasses, etc.

Objects of every class and their descriptions are simultaneously numbered in such a way that objects of zero rank are numbered first (lowest degree of classification), then of the first rank, etc.

In our case we are dealing with a small number of types of objects (vessels, ports, cargoes, clientele, adjacent forms of transport, etc.).

One of the labor-consuming problems in this plan is classification of objects (elements) of the information system and also the establishing of characteristics of objects and their classification.

Objects and characteristics must be classified proceeding from conveniences of solution of problems, and retrieval of necessary information. The set containing the greatest number of characteristics is the set of information about vessels. Vessels, which will be processed by the ports of a given steamship company, i.e., enter the sphere of its control, will be both domestic and foreign. In turn, every group is divided into cargo, cargo-passenger, and passenger vessels. Every class is divided into smaller subclasses. The smallest unit of classification is a group of vessels of a definite type, for example, of the type "Poltava," "Krasnograd," etc., having many identical characteristics. These types take into consideration design features and adjustability to transport of definite cargoes. Every vessel in the zero class should possess a certain formal criterion (index), which takes this classification into consideration. Many of the problems of operational regulation of the work of the fleet are solved for interchangeable hawsers and vessels. Therefore, in the process of solving a problem from the formal criterion (index) the machine should determine what vessels can be used in every concrete case.

The second important group of objects and a rather large one is ports. In this group there are all ports, in which the vessels of a given basin are processed: domestic ports of the given basin and other basins, and also foreign ports. Every group of ports is distinguished by composition of necessary information.

If classification of vessels and ports is rather simple, then classification of other objects and sources of information of the system and formalization of their characteristics have difficulties. This pertains especially to loads. For this purpose it is necessary to develop anew a single product-list of cargoes, which is needed because at present different product-lists exist: one for determination of tariffs on transport of cargoes in export-import and cabotage, others for standardization of loading and unloading of works, and still others for capture of cargo collection. All these product-lists have noncoincident designations of cargoes, and to store in the machine all product-lists is inexpedient. Therefore, a new product-list of cargoes should consider in the form of a characteristic all necessary indices.

Every information table consists of three parts:

cap of objects,

cap of characteristics,

table of values.

The principle of composition of an information table is well-known [3]. In the beginning designations of all objects of class M are recorded and numbered. Then there are recorded designations of all characteristics (criteria) describing the state of the objects of the given class in a certain sequence (a procession of criteria is composed), and they are numbered.

In the first line of the information table criteria are recorded in the order of their numbers. In the second line there are recorded under every designation the values of the given characteristic belonging to object No. 1. In the third line there are placed characteristics for object No. 2, etc.

The cap of objects is a vertical cap since every assignment of the object corresponds to a line of information table. The cap of characteristics is horizontal because every assignment of characteristic corresponds to its concrete value for every object.

In caps of objects, besides ideas expressing the name of concrete objects or their coded equivalent, there are given links of entry of classes of some ideas into classes of other ideas.

It is necessary to note that characteristics can also be classified, i.e., characteristics can also be divided into classes, subclasses, etc., by ranks. For example, such a characteristic of a vessel of the first rank as "coordinate" is divided into two characteristics of zero rank – "latitude" and "longitude." Therefore, in caps of characteristics also, besides the name of characteristics, there are given links of entry of characteristics of junior ranks into classes of senior ranks and their links with the matrix. This is done for convenience of retrieval and delivery of a reference. During retrieval of information there should be assigned the rank of the idea of the object or characteristics by which the necessary values are selected.

In one information table one should place data about objects the majority of the characteristics of which have identical names, and there must also be considered the convenience of retrieval of information for solution of problems. Thus, for example, it is expedient to have separate information tables for oil, passenger, and dry cargo fleets.

The information table for cargo vessels of the dry cargo fleet includes technical characteristics of vessels (carrying capacity, load-carrying capacity, speed of movement, type of engine, number of holds, hatches, etc.), normative data (norms of consumption of water, fuel on water and on stand, primecost of vessel, norm of consumption of currency, etc.), plan of work of vessel in current trip (scheduled assignment), location of vessel and form of fulfilled operation by last operational data, fulfillment of scheduled assignment, data about fulfillment of preceding trips (time of trip, designation and quantity of transported load, material and financial consumptions, sum of prize, etc.). The vertical cap of this table is of the third rank and contains around five hundred designations of characteristics. The volume of the horizontal cap depends on the number of vessels of steam navigation and can also contain more than one hundred designations.

There may also be cases in which retrieval detects not one of the ideas of the object or characteristic. In this case there are generated special criteria equivalent to the expressions "there is no assigned idea in the cap," "there are no ideas of zero rank subordinated to the assigned idea in the cap." If for description of a certain object a certain characteristic is useless, the criterion "does not have meaning" is introduced, which is placed on the place of concrete value of the characteristic. Furthermore, there is introduced the official criterion "there is no information," which is used if information about the value of the given characteristic for any reason did not enter the information system.

Thus, the information system is in the form of information tables. The finding of the concrete value of a characteristic is determined by the information retrieval algorithm for which there must be assigned certain initial data: number of information table, number of line of information table (name of object) and kind of concrete characteristic of object corresponding to name (number of column).

Algorithms of solution of concrete problems developed at present do not provide for machine preparation of initial data and, therefore, require large expenditures. During automation of control of the transport process it is necessary to liberate maintenance personnel from labor-consuming work on data processing. Subsequent organization of solution of the problem in a system of information provision is possible.

Algorithms of solution of problems of control of the transport process will be developed beforehand, and programs of their solution will be presented in the form of a library of standard subroutines.

The solution of every problem is divided into two stages:

– preparation of initial data;

– solution of the problem according to obtained data.

Therefore, programs of preparation of initial data have to be separated into a definite group of problems. In each program of preparation of initial data there participate information-retrieval algorithms which depend on the organization of the information depository.

For simplification of distribution of information about the transport process a certain assumption is made in our case: let us keep in mind that the number of system elements is constant. Furthermore, let us take into consideration that the number of characteristics necessary for description of the properties of every object is also known.

Knowing the limits of change of the contents of the characteristics, it is possible to establish the volume of information which will describe the state of the system in a certain interval of time.

An important question of investigation is development of a method (language) of presentation of information inside the machine, ideas of caps of objects, and characteristics and their values. Closely connected with this is construction of classificational links between objects and characteristics and also application of methods of distribution of information in memory units.

The main part of the information will be stored on magnetic tape in linear form. In this sense the information table should be placed line by line or column by column. In our case it is expedient to place information line by line, forming separate arrays for every object.

To ensure compact recording of information in memory there are used several special procedures. The simplest is the position principle of distribution. In this case for every object there is assigned a strictly defined volume of memory for storage of information about it. In the process of work there will be changed the contents of cells storing variable information, and the character of information and cells assigned for their storage remain constant. Under the value of the characteristic there is assigned a constant number of binary digits corresponding to its maximum value.

The main drawback of such a method is uneconomic use of memory when there are many empty values of characteristics. Other methods of information consider nonzero values of characteristics, which are recorded sequentially one after the other. Upon the finding of information there are used various kinds of logical scales, which are at the beginning of the array of characteristics for a certain object. These scales consider the presence of characteristics and their dimension.

Selection of a method of distribution of information tables can be solved after their construction and solution of the problem of presentation of information.

Questions of development of methods of data processing are not less complicated. For purposes of transmission and processing, economic information, as a rule, is grouped in reports. At present there have been developed basic principles of processing of data on control of naval transport. As a basis of developed algorithms there is assumed the principle of clear separation of algorithms of accumulation of information from algorithms of delivery.

Algorithms of accumulation of information are algorithms of input and restoration of information, and algorithms of delivery are algorithms of obtaining answers to inquiries and preparation of initial data for solution of individual problems. A method of construction of algorithms and programs of processing of operational information about the work of the naval fleet has been developed by Tseytin, the senior scientific colleague of the A. A. Zhdanov Leningrad State University. As a basis of algorithms of data processing there is assumed treatment of separate elementary reports containing one or more characteristics of objects.

The first variant of IPS for automated control of the work of a fleet will be created on the basis of the Experimental Computer Center of the Naval Fleet in Baltic Steam Navigation.

Bibliography

1. Plyatsidevskiy I. Informatsionnyye sistemy v tekhnike i ekonomike (Information systems in technology and economics). Moskovskiy rabochiy, 1966.

2. Abramov O. A. and Batrakov V. I. Elektronnyye tsifrovyye mashiny i snabzheniye voysk (Electronic digital computers and supply of troops). M., Voyenizdat, 1964.

3. Krinitskiy N. A. Tablitsy ob"yektno-kharakteristicheskiye. Promyshlennaya elektronika i avtomatizatsiya proizvodstva (Object-characteristic tables. Industrial electronics and automation of production). Tom III, M., Sovetskaya entsiklopediya, 1965.

EXPERIMENTAL INVESTIGATIONS OF COMPARATIVE EFFECTIVENESS OF MANUAL AND MECHANIZED IPS IN THE N. K. KRUPSKAYA LENINGRAD STATE INSTITUTE OF CULTURE

A. V. Sokolov, D. I. Byumenau, R. F. Grinina,
and A. M. Sorkin

The present work deals generally with the problems, method, and results of three interconnected experiments conducted in 1964-1966 in N. K. Krupskaya Leningrad State Institute (LGIK) under the general names of "Lastochka," "Estafeta," and "Ruduga." Experiments were prepared and conducted by a 10-man initiative group. The practical target of investigations consisted in obtaining initial data for determination of rational fields of application of manual and mechanized information retrieval.

A. Problems of Experiment

Effectiveness is a relative concept. The conclusion that an IPS is more effective can be made only on the basis of its comparison with other IPS. In LGIK experiments manual card files were compared to mechanized IPS of the descriptor type. Effectiveness of IPS was estimated according to the following indices, determining both the final useful effect and the expenditures, made for its achievement:

1. The quality of work determined by information losses and information noise.

2. Labor-input of creation.

3. High speed operation.

The main criterion of effectiveness was considered to be information losses and the main attention was paid to exposure of regularities affecting this index. Experiments were conducted on the basis of three different collections of documents on library matter and scientific information with the use of both inquiries formulated by the experimenters themselves, and inquiries assigned by information users.

Manual IPS were investigated in two directions:

a) appraisal of subject and systematic catalogs[1] accepted in library and information practice, called "traditional IPS" from now on;

b) exposure of promising ways of departure from traditional practice within the bounds of manual retrieval.

Descriptor IPS participating in experiments were created in accordance with rules generalized in [1] and can be considered by typical representatives of the given class of IPS. Descriptor language was represented in the form of a thesaurus on library matter and scientific information developed at the N. K. Krupskaya LGIK.

Appraisal of effectiveness of traditional IPS as compared to descriptor systems is the purpose of the works "Lastochka" and "Estafeta." Analysis of catalogs and card files existing in information services and libraries permitted to set that the following peculiarities are inherent to them as IPS:

1. As an IPYa there comes forward an a priori assigned

[1] In this work we do not distinguish between the terms "card file" and "catalog."

hierarchical diagram of classification (for systematic catalogs) or empirical list of subject heading with cross references (for subject catalogs).

2. There is used a criterion of semantic conformity "on inclusion," according to which in delivery there are included documents carrying indices equal or subordinated with respect to the indices of the retrieval pattern of the inquiry.

3. A document yields an average of 1-2 independent indices and, accordingly, the norm of duplication is not more than 1.5 cards per document.

4. Realization in the form of a manual-retrieval card file.

Traditional card files for LGIK experiments were developed by several bibliographers with a good knowledge of accepted library practice. Their problem was to create bibliographical card files satisfying contemporary requirements for such card files. There were no limitations on the finishing of available classifications, depth and detailedness of indexing and degree of duplicating of cards. Bibliographers worked independently of each other. It is significant that card files obtained as a result corresponded to the above-formulated "traditional norm," which once again confirmed the justice of these norms.

Table 1 gives the conditions of carrying out the experiments "Lastochka" and "Estafeta" and the obtained results. As can be seen from the table, the difference in indices of information losses of traditional card files and descriptor IPS clearly exceeds the bounds of possible inaccuracy of experiment, therefore a conclusion can be drawn concerning the superiority of typical descriptor IPS over typical traditional card files in the sense of the quality of retrieval. This conclusion is important in itself, but is still insufficient for determination of rational fields of application of manual and mechanized retrieval technology. There remained open the question of the possibility of improvement of manual IPS by way

Table 1.

.Information-retrieval systems	Condition of experiments			Indices of effectiveness			
	degree of duplication of cards	number of inquiries	file of documents	loss of information, %	information noise, %	retrieval time of 1 inquiry, min	processing time of 1 document, min
"Lastochka"		40	1600				
UDC (traditional)	1,2			48	60	7,35	2,41
Descriptor IPS	—			9.3	22	5,8	9,42
"Estafeta"		90	2300				
Subject catalog (traditional)	1,37			46	4,5	—	—
Descriptor IPS	—			14	15,5	—	—
"Raduga (first stage)		100	800				
BBK (traditional)	1,3			56.3	54	4	—
SDK (untraditional)	3,5			35.3	15	3,2	—
Subject catalog (untraditional)	3,8			16.6	24	7	—
Descriptor IPS	—			17,7	16,6	2,5	—
"Raduga" (second stage)		100	1600				
BBK (traditional)	1,31			56,2	65,5	7,3	2,1
Subject catalog (traditional)	1,27			64.8	38	5,1	2,0
SDK (untraditional)	3,37			39,2	35	4,8	5,4
Subject catalog (untraditional)	3,7			14,9	26	8,5	6,4
Descriptor IPS	—			7,0	30	5,0	9,0

of deviation from certain "traditional norms." To determine ways of increasing the quality of work of traditional IPS we will analyze sources of information losses and information noise inherent to them and ways of removing them:

1. Apriority of classification diagrams utilized for construction of systematic catalogs. Preassigned classifications are in no way oriented to the available file of documents; ideas and links essential for description of the documents of a given file are always absent in them. For the purpose of compensation of apriority of classification diagrams there is practiced their finishing in the process of exploitation. As a result the a priori assigned diagram approaches a classification diagram developed empirically, proceeding from subjects of given array, and theoretically in limit should merge with it. During the carrying out of the experiment it is possible to exclude the influence of apriority by way of use of empirically composed classification.

2. Insufficient depth and detailedness of indexing, due to which the information contents of the document are not completely reflected in the retrieval pattern of the inquiry. The traditional method, as shown above, establishes a norm of duplicating not more than 1.5 cards per document, thereby limiting the possibilities of indexing. In principle this limitation can be removed without changing the specific character of the system.

3. The linearity of indices of manual catalogs manifested in the fact that heuristic functions are inherent to only the left part of the complicated index; nevertheless, the remaining components (in particular, model subheadings and determinants) do not fulfill heuristic functions. Removal of linearity of indices within the limits of manual card files is possible by way of duplicating of cards, introducing as many cards per document as there are elements in the retrieval pattern of the document.

Possibility of decreasing information losses within the bounds of manual IPS are covered by "Raduga." The task of "Raduga" was to construct manual-retrieval card files in which the above-mentioned sources of information losses are removed. The problem of "Raduga" corresponding to the second direction of investigations of manual IPS was formulated in the following way: whether there exist objective causes preventing the obtaining of identical completeness of delivery of information in manual card files and in model descriptor IPS during equal conditions of processing and retrieval of documents? In order to check the authenticity of estimated data obtained in "Lastochka" and "Estafeta" the "Raduga" program provided for the creation of traditional subject and systematic catalogs. A traditional systematic catalog was organized according to new Soviet Library-Bibliographic Classification (BBK).

In contrast to preceding experiments, where IPS were constructed independently of each other, in the case of "Raduga" it was required to provide coordination of untraditional manual card files with descriptor IPS contrasted to it in order to exclude the influence of subjectivity of operators of these systems. Such a measure is necessary, inasmuch as any IPS belongs to the class of "man-machine" systems and, therefore, the effectiveness of its fulfillment of functions depends both on the objective possibilities of the IPS as a machine and the subjective qualities of the person of the operator working with it. If one does not provide for measures for exception or, in the extreme case, equalizing of the influence of subjective factors, then the study of objective characteristics of the system becomes impossible. In the "Raduga" experiment the following measures were taken to compensate for subjectivity:

a) information-retrieval languages were equalized with respect to semantic force;[1]

b) description of information contents of documents (indexing) was carried out with an identical degree of depth and detailedness;

c) programs of information retrieval on inquiries were coordinated;

d) retrieval results were evaluated on the basis of single criteria.

In order to satisfy the first condition, for an untraditional systematic catalog there was developed a Special Decimal Classification (SDK) of literature on library science, bibliography, and scientific information. The SDK was constructed according to the type of traditional "enumerating" diagrams: structurally it presents a single hierarchical "tree" of ideas, there is applied decimal notation and a scanned system of model divisions (determinants) and provision is made for the possibility of formation of complicated indices with the help of "colon" and "plus" signs. A peculiarity of SDK is that its glossary matches the glossary of an empirically

[1] Semantic force means the possibility of describing phenomena by means of a given language. Semantic force determines possible depth and detailedness of indexing.

compiled thesaurus. Therefore, one may assume that SDK is an empirically created classification. Comparing SDK with an a priori assigned BBK, it is impossible not to note the considerably great complexity of SDK tables. Into SDK there went 1109 different indices, whereas in the BBK card file only 482 were used. The problem of equalizing the semantic force of the descriptor language and the language of subject headings was solved in the process of formulation of subject headings by way of coordination of them with the retrieval pattern of the document in descriptor language. The subject headings of the untraditional catalog included frequently three or more subheadings, whereas, in the traditional subject catalog the structure of the headings was much simpler (as a rule, title and subtitle).

For compensation of subjectivity of indexers in untraditional IPS and in descriptor IPS there was carried out standardization of the retrieval pattern of the document. Standardization consisted in the fact that the retrieval patterns of documents composed in various IPYa included the same idea. Thus, there was ensured identical depth and detailedness of indexing. Standardization does not ensure absolutely correct indexing (this is practically impossible); the purpose of standardization is to provide an identical level of errors and inaccuracies in all untraditional IPS.

On every independent SDK index, besides determinants, there was given a separate card in the card file. In exactly the same way on every significant word of a subject heading, with the exception of model subheadings, corresponding to SDK determinants, an additional card is started. As a result, as Table 1 shows, the degree of duplication in untraditional catalogs considerably exceeded the usual library norms. In descriptor IPS there were used during indexing an average of 5.65 descriptors per document.

To get rid of subjectivity in understanding inquiries and composing the retrieval program the retrieval patterns of inquiries of untraditional manual card files and the descriptor system were intercoordinated in such a way as to achieve standardization of the retrieval pattern of the inquiry in exactly the same way as standardization of the retrieval pattern of the document was provided.

The relevance of documents was evaluated by a competent commission on the basis of single principles shown below. To eliminate tendentiousness in evaluation of relevance, experiments were organized in such a way that members of the commission did not know what document was issued by one system or another.

Thanks to the above-described measures there were excluded two sources of losses in traditional IPS: apriority of classification diagram and insufficient depth of indexing. Partially there was compensated also linearity of indexation of manual systems. It is true that compensation of linearity was not complete since heuristic functions were nob given to SDK determinants and model subject subheadings. Calculation showed that if this was done, then the volume of untraditional catalogs would be increased 2-3 more times, and the degree of duplication would reach 6-7 cards per document.

B. Method of Experimenting

During the carrying out of experimental IPS investigations wide propagation was obtained by the method accepted during realization of the Cranfield project [2]. By this method every inquiry participating in the experiment is formulated on the basis of a document-source arbitrarily selected from the file in such a way that the document-source completely answers it. The number of the document-source is reported together with the inquiry to the person doing the retrieving. Retrieval on the inquiry continues until the IPS gives out the document-source or permissible retrieval variants are exhausted. In the first case retrieval is considered successful, and in the second unsuccessful. The total percentage of unsuccessful retrievals determines information losses.

The advantages of the Cranfield method are simplicity and convenience of experimentation. In the opinion of the authors of the project, the method proposed by them permits excluding the complicated question of evaluation of relevance of documents to the given inquiry, since the number of the document-source is known beforehand. At the same time there are doubts with respect to the reliability of this method [3, 4, 5]:

1. It is possible to assume that the probability of non-issuance of a document-source depends on information losses inherent to IPS, but this assumption requires proof.

2. Inasmuch as every inquiry corresponds to a document-source the connection of which with the inquiry is not stipulated by the method, instead of the problem of relevance between the inquiry and documents as a result of retrieval there appears the problem of relevance between the document-source and the inquiry made on the basis of it. For strict carrying out of the experiment it is necessary to formulate a clear criterion of conformity between the contents of the document-source and the inquiry. To do this is as difficult as to stipulate the conditions of relevance of a random document to a given inquiry. Thus, despite affirmation of the authors, the Cranfield method does not exclude the problem of relevance.

3. The impossibility of use of "real" inquiries of users is in no way connected with experimental collection of documents.

4. The Cranfield method does not permit calculating IPS information noise.

In the N. K. Krupskaya LGIK there was developed a more exact, in our opinion, and more complicated method of experimentation intended for comparative investigations of two or more IPS. According to the LGIK method, information losses and information noise for n inquiries are calculated by direct means by formulas (1) and (2):

$$L=\frac{1}{n}\sum_{i=1}^{n}\frac{S_{\Sigma_i}-S_i}{S_{\Sigma_i}}\cdot 100\%, \qquad (1)$$

$$N=\frac{1}{n}\sum_{i=1}^{n}\frac{P_i-S_i}{P_i}\cdot 100\%, \qquad (2)$$

where L is information loss; N is information noise; S_{Σ_i} is the total number of relevant documents in an array for the i-th inquiry; S_i is the issued number of relevant documents for the i-th inquiry; and

P_i is the total number of documents issued in answer to the i-th inquiry.

The problem of relevance is solved on the basis of the following considerations. Among documents issued by an IPS in answer to an inquiry there are always documents which can be confidently recognized as relevant or, conversely, irrelevant to a given inquiry. Authenticity of determination sharply increases if it is carried out jointly by a special commission. Uncertainty in judgement of relevances spreads only to a certain, as a rule, small number of documents. By way of these documents there is created inaccuracy of experiment, but the possibility of realizing the experiment is not negated by the presence of "indefinite" documents. Practice showed that during the joint method of determination of relevance the shown "uncertainty" is successfully solved. In the LGIK there have been established the following leading rules for the evaluation of the relevance of documents to an inquiry:

1. Relevance must be evaluated on the basis of the source used during indexing. If indexing is carried out according to annotation or abstract, then accessing the primary source or conjecturing its contents is not allowed.

2. Relevance is determined irrespectively of the reader's assignment of the document (theoretical article, patent, popular work), if the reader's assignment is not stipulated in the request.

3. If the document concerns a less general idea than the idea assigned in the inquiry, then the document is considered relevant; if, however, a more general idea appears in the document, then it is not recognized as relevant.

The described method can determine S_i for every inquiry and every IPS. As S_{Σ_i} is the total useful delivery of all IPS. Such an assumption is based on the fact that delivery of systems, as practice shows, is additional one to another, i.e., relevant documents not issued by one system are issued by another, and vice versa. The more systems participate in the experiment, the nearer sum S_i approaches the hypothetical value of S_{Σ_i} appearing in formula (1). It must be borne in mind that an absolutely accurate determination of information losses is not necessary to come to a conclusion about the superiority of one IPS over another; it is enough to be convinced that one IPS issues a greater number of relevant documents than another, i.e., to obtain the difference value of information losses of investigated IPS.

The LGIK method has, in our opinion, the following merits:

a) direct calculation of information losses and information noise increases authenticity of obtained results;

b) the possibility of using "real" inquiries of users;

c) the possibility of calculating information noise.

The deficiencies of the method are the necessity of the participation in the experiment of a minimum of two IPS; and the relative accuracy of determination of S_{Σ_i}.

C. Results of Experiments and Conclusions

Table 1 gives actual data of LGIK experiments. Values of indices of effectiveness of traditional IPS obtained in different experiments closely match, in spite of the fact that in every experiment these IPS were composed by different specialists on the basis of different literature bases and different sets of inquiries were used. Consequently, these indices are of an objective nature and do not depend on the subjectivity of the operator or the conditions of the experiment. In particular, it is possible to deduce the following mean information losses and labor consumption of input of one document in traditional and descriptor IPS (Table 2).

Table 2.

IPS	Information losses, %	Information noise, %	Input time of document, min
Systematic catalogs (traditional)	53	60	2.26
Subject catalogs (traditional)	55	21	2.0
Descriptor IPS	12	21	9.21

Table 2 data show that the quality of work and labor consumption of creating traditional systematic and subject catalogs are on one level. Information losses in typical traditional catalogs are 40% higher than in typical descriptor IPS, but the time spent processing documents in the first is four times less. Thus, the first problem of the experiments (evaluation of the effectiveness of traditional retrieval technology) can be considered carried out.

In order to answer the question of whether there are objective causes preventing the obtaining of an identical level of information losses in manual card files and descriptor IPS (the second problem of investigation), let us analyze the causes of appearance losses in untraditional catalogs of the "Raduga" experiment (second stage). The experiment showed that in these catalogs there are differences in losses of information from descriptor IPS, in spite of steps taken to compensate for the subjectivity of operators of these IPS. The following sources of information losses were exposed (Table 3).

Technical errors were headpiece of cards or their omission during retrieval, incorrect writing of index on card, absence of index entering standard retrieval pattern, and incorrect puncture or error during readout of punched cards.

Table 3.

Sources of losses	Information losses, %		
	SDK	subject catalog (untraditional)	descriptor IPS
Technical errors	11.9	1.2	3.9
Subjectivity of indexing	2.8	2.7	2.8
IPYa defects	4.3	2.1	0.3
Linearity of indices	20.2	8.9	—
All	39.2	14.9	7.0

Subjective errors of indexing were defects of the standard retrieval pattern (insufficient depth or detailedness of indexing). IPYa defects consisted in absence of basic links between IPYa elements useful for retrieval. Losses due to linearity of indices of manual IPS are caused by the impossibility of retrieval according to model subheadings and determinants, since these language elements did not give heuristic functions.

It is obvious that there will always be technical errors in systems of the "man-machine" type including a person as one of the sections. In exactly the same way there are removed errors connected with the subjectivity of indexers, although thanks to standardization of retrieval patterns these errors can be brought to one level. The inevitability of the presence of the two sources of information losses shown proves the impossibility of practical construction of real IPS having zero information losses.

With the specific character of IPS there are connected losses caused by IPYa defects and linearity of indices. In descriptor IPS the influence of the first source of information losses is insignificant, and the second source is absent. Inasmuch as basic links were established by IPS operators on the basis of erudition and intuition, then omissions and errors in exposing them are inevitable in any IPS. But the specific character of the language of manual systems is that in them the probability of appearance of defects is greater than in descriptor IPS. With increase of depth and detailedness of indexing there is quickly complicated the structure of the hierarchical diagram of classification and the dictionary of subject headings. At the same time achievement of such quality of indexing in descriptor IPS does not require much complication of the thesaurus, thanks to the fact that every word of descriptor language possesses heuristic functions. It is not difficult to see that "simplicity" of descriptor language is the result of the basic distinctive peculiarity of systems of coordinate indexing – the possibility of retrieving according to any IPYa elements and their combinations. Manual means of realization do not allow such a possibility. Thus, in the case of manual IPS there is obtained a closed circle: to decrease losses of information it is necessary to increase depth and detailedness of indexing, which cannot be done without development of complicated IPYa possessing great semantic force. In turn, IPYa complication inevitably leads to errors and omissions during its creation and use, which involves information loss.

Linearity of indexation in manual catalogs, as shown above, can be completely excluded by way of increasing degree of duplicating of cards. In the "Raduga" experiment we refused formation of independent indices on the base of determinants of classification and model subheading since this would contradict the main purpose of these elements – to serve as auxiliary means for more accurately defining of basic indices IPYa. But in principle such measure is not excluded.

Returning to the main question posed in the "Raduga" experiment we can ascertain that there were not revealed theoretically irremovable obstacles to bringing the level of information losses in manual card files to the level of information losses in descriptor IPS if initial data are equal. However, practical achievement of this level is hampered by increasing complication of manual systems. In systematic catalogs a practically irremovable source of information losses is apriority of classification since introduction of empirical diagrams apparently is impossible.

Conclusions

1. It is practically impossible to construct bibliographic IPS possessing an "ideal" quality of work, i.e., zero information losses.

2. Traditional IPS having comparatively simple IPYa structure and using small depth and detailedness of indexing independently of the principle of organization of the card file — subject or systematic — have approximately 40% higher information losses than model descriptor IPS.

3. The tendency to compensate sources of information losses of traditional IPS within the bounds of manual retrieval leads to increase in the volume of card files and the labor consumption of their creation and to considerable complication of IPYa structure, which is a potential source of information losses.

4. Descriptor IPS in principle must provide a minimum level of information losses but realization of such systems takes much more work than realization of manual IPS.

Actual data obtained by us can be used in further investigations aimed at determining the rational fields of application of various IPS. In this we see the basic meaning of experiments conducted at the N. K. Krupskaya LGIK.

Bibliography

1. Mikhaylov A. I., Chernyy A. I., Gilyarevskiy R. S. Osnovy nauchnoy informatsii (Bases of scientific information). M., "Nauka", 1965.

2. Cleverdon C. W. Report on the testing and analysis of an investigation into the comparative efficiency of indexing systems. Aslib Cranfield research project, 1962.

3. Taube M. A note on the pseudomathematics of relevance. Amer. Docum., v. 16, 1966, № 2, 69—72.

4. Swanson D. R. The evidence underlying the Cranfield results. Library Quarterly. v. 35, 1965, № 1, 1—20.

5. Rees A. M. The Cleverdon — WRU experiment search results. In: Information retrieval in action W. R. U., 1963, 93—99.

INFORMATION RETRIEVAL SYSTEMS

R. I. Trukhayev and V. V. Khomenyuk

A description of an information retrieval system, its functional stages and structure is given.

1. Information retrieval systems are one of the means and component parts of an organ for making a decision (by which is understood a group of people, one person, a technical device or a complex of technical devices, etc.) during the study and explanation of the character of functioning of different objects (processes, phenomena and others) for the purpose of output of a solution on control of this object, and also for the purpose of accounting for the possible influences from the side of this object (in case it actively or passively counteracts).

Functioning of an object is examined as a sequence of transitions (in time) of an object from one position x in n-dimensional metric space into another possible position $y \in R$.

Two basic types of information retrieval problems are possible (according to the character of the goal searched) depending on what is required: to separate an object with certain criteria from a great number of analogous objects or to determine the state of the object from a certain large number G of possible states.

By position of the object is understood, in the first case, the vector of characteristic criteria of the object, in the second case, the vector of spatial coordinates.

Information retrieval systems are intended for determination of the position of an object. What does application of information retrieval systems give?

Information retrieval systems permit determining the position of an object in time with a certain degree of definitiveness depending on the time of functioning, construction and other parameters of the information retrieval system. Obtained information about the position of an object permits the organ making the decision to carry out its functions: to make one or another decision in accordance with goals of its work. For example, an information retrieval system should find the code of a book from a catalog and (if this is necessary) find the book and issue it to the reader.

Further on we will briefly examine the principles of functioning, capabilities, structure and criteria of work of information retrieval systems.

2. The essence of functioning of an information retrieval system consists in determination (finding) the position of an object on the basis of a certain model of functioning of the object, data obtained during search, observation and processing of available informations about the object.

It is necessary to note that information retrieval systems should, as a rule, carry out a purposeful or goal-directed process of retrival of information about the position of an object, since there are different criteria and limitations on parameters of functioning of information retrieval systems.

Functioning of information retrieval systems can be split into the following stages (see Fig. 1):

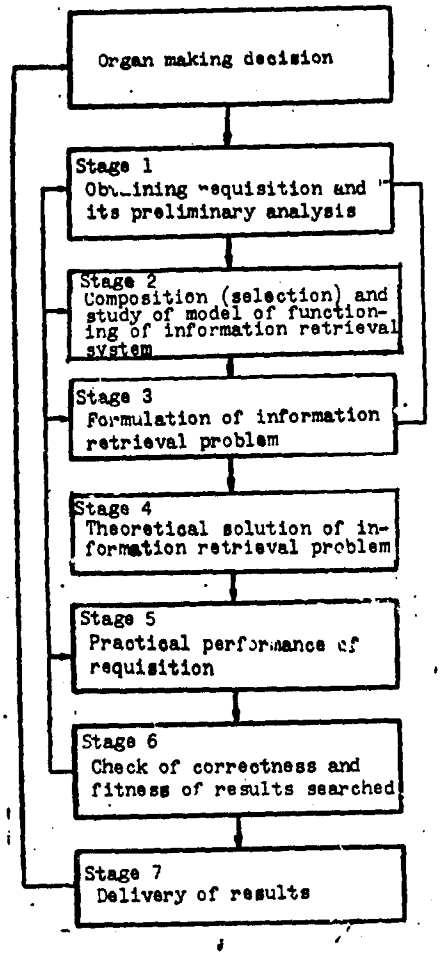

Organ making decision
Stage 1
Obt‿ining ˇequisition and
its preliminary analysis
Stage 2
Composition (selection) and
study of model of function-
ing of information retrieval
system
Stage 3
Formulation of information
retrieval problem
Stage 4
Theoretical solution of in-
formation retrieval problem
Stage 5
Practical performance of
requisition
Stage 6
Check of correctness and
fitness of results searched
Stage 7
Delivery of results

Fig. 1. Stages of functioning of an information retrieval system.

1) obtaining a requisition or order (from the organ making the decision) and its preliminary analysis;

2) composition (or selection) and study of qualitative and quantitative models of functioning of the object;

3) formulation of the information retrieval problem;

4) theoretical solution of the information retrieval problem posed;

5) practical performance of search;

6) check of correctness and fitness of data obtained by the information retrieval system;

7) readout and results to organ making the decision.

The first stage consists in obtaining a requisition, from the organ making the decision, formulated according to set standard rules (in the form of a formula, oral assignment and so forth). In the requisition are indicated:

a) criteria of the object separating it from analogous objects, or certain information about its position accordingly for problems of the first and second types;

b) requirement about finding the actual object or requirement about finding the position of the object;

c) criteria and limitation on the character of performance of the requisition (time of performance, required accuracy and volume of information, etc.);

d) form of delivery of the information obtained by the information retrieval system to the organ making the decision.

Preliminary analysis consists of the fact that a requisition is formulated in accordance with standard rules, and in preliminary appraisal of the possibility of use of the obtained requisition in the information retrieval system.

The second stage consists of the fact that on the basis of study of the actual object and other indirect information about the object a qualitative or quantitative model of functioning of the object is composed (selected).

The third stage consists in presenting the inf .mation retrieval problem on the basis of:

a) given requirements from the requisition;

b) model of functioning of the object;

c) capability of the information retrieval system;

d) criteria of functioning of the information retrieval system.

Mathematically, information retrieval problems are formulated as problems of finding solutions to problems of mathematical logic, problems of mathematical statistics and the theory of statistical solutions, systems of equations and inequalities, on the one hand, or solution of extreme problems in the presence of limitations, etc., on the other hand.

The fourth stage consists in development and application of methods for solution of a posed information retrieval problem, and also in a check of correctness of obtained (on the basis of these methods) a theoretical solutions of an information retrieval problem.

The fifth stage consists in practical performance, by an information retrieval system, of a received requisition or order on the basis of solution of the information retrieval problem.

The sixth and seventh stages consist in a check of correctness and fitness of data obtained by an information retrieval system and delivery of the obtained results to the organ making the decision.

3. In accordance with stages of functioning, an information retrieval system has the following structure (see Fig. 2).

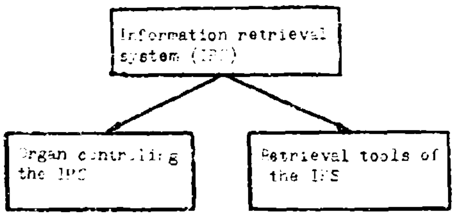

Information retrieval system (IPS)
Organ controlling the IPS
Retrieval tools of the IPS

Fig. 2. Structure of information retrieval system.

An information retrieval system consists of retrieval tools and the organ of control of the system.

The tools can constitute different technical devices, constructions and attachment, and also complexes of technical devices together with personnel. Not stopping for detail on principles of devices and work of retrieval tools, let us give an idea about the retrieval unit, the space searched, density of distribution of retrieval equipment and its volume.

Information retrieval systems consist of a set of a varying quantity of various types of retrieval equipment.

A retrieval unit of a certain type is that quantity of equipment of this type which has its own control system and is intended for production of a defined character of information about an object. A retrieval unit of a system can be a certain technical device, a person or group of people, a complex of technical devices with personnel, etc., which are intended for production of information about the object. The concept of a retrieval unit is conditional, just as any other concept of unit of weight, length, area, etc.

The space searched by the tools of an information retrieval system is either the space and time coordinates of the tool, or the character and quantity of information about the object (obtained from the retrieval tool), or a combination of them.

Density of distribution of the equipment of an information retrieval system is defined as the quantity of retrieval units of different types belonging to a great number in space searched unit measure.

The volume of retrieval equipment of an information retrieval system is defined as the total quantity of retrieval units of different types distributed in the whole space searched. It is clear that the volume is determined by unification with the sum of volumes of uniform equipment. Here under volume of uniform equipment is understood the quantity of units of one type distributed in the space searched.

Control of an information retrieval system constitutes a technical device or group of people, a complex of technical devices with personnel intended for:

a) development of distribution of retrieval tools;

b) control;

c) delivery to organ making a decision.

The structure of control of an information retrieval system is represented in Fig. 3.

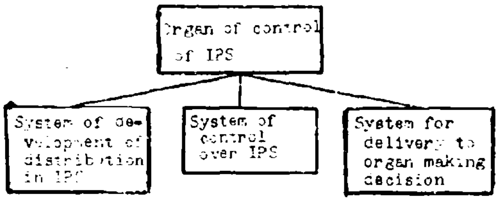

Organ of control of IPS
System of development of distribution in IPS
System of control over IPS
System for delivery to organ making decision

Fig. 3. Structure of control of information retrieval system.

According to stages of functioning of an information retrieval system, the system for development of distribution means of retrieval is fulfilled by stages 1-4 and 6, control of the system by the 5th stage, and the system of delivery to the organ making the decision by the 7th stage.

The system of development of distribution of retrieval of an information retrieval system consists of 6 blocks:

a) block for reception of requisition and its preliminary analysis;

b) block of composition (selection) and study of model of functioning of object;

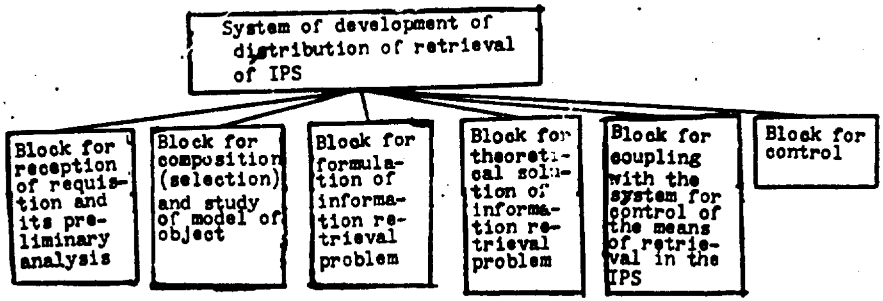
c) block of formulation of information retrieval problem;

d) block of theoretical solution of information retrieval problem;

e) block for coupling with the system of control of means of retrieval of an information retrieval system;

f) block of control.

The structure of the system for development of distribution of means of retrieval of a system is represented in Fig. 4.



System of development of distribution of retrieval of IPS
Block for reception of requisition and its preliminary analysis
Block for composition (selection) and study of model of object
Block for formulation of information retrieval problem
Block for theoretical solution of information retrieval problem
Block for coupling with the system for control of the means of retrieval in the IPS
Block for control

Fig. 4. Structure of system of development of distribution of means of retrieval of an information retrieval system.

4. The information retrieval problem is that of determination of distribution of means of retrieval of the system in the space searched taking limitations and criteria into account.

Limitations and criteria of the problem arise from requirements of the requisition from the organ making the decision, the accepted model of functioning of the object, capabilities and requirements of work of the system, i.e., parameters of the control system and means of retrieval of the system.

The character of limitations and criteria of the problem can be represented in a definite form, in a probabilistic form (assigned probabilistic characteristics) and in an indeterminate form in accordance with the form of the requisition, the composition (selection) of the model of functioning of the object, representation of work of control and parameters of means of retrieval of the system. For example, the requisition can show all, or perhaps not all data for its fulfillment. In addition, the model of the object can be composed in a definite probabilistic or indefinite couplings (equations, inequalities, functional relationships, etc.). Furthermore, knowledge of parameters of the information retrieval system can also be presented in a determined accidental or indeterminate form.

It is necessary to note that in the first stage of posing and solution of an information retrieval problem a great degree of indeterminateness (i.e., chance and uncertainty) is introduced by parameters of the model of functioning of the object.

In practice, limitations and criteria are expressed in different form; for example, criteria and limitations of work of an information retrieval system can be:

a) readiness of the system to fulfill the requisition;

b) probability of fulfillment of the requisition per unit time;

c) probability of obtaining an assigned volume and accuracy of information about the object per unit time;

d) cost of fulfillment of the requisition per unit time;

e) time of fulfillment of the requisition;

f) volume of means of retrieval;

g) degree of automation of the system;

h) simplicity, adaptability (fitness for performance of a large number of mixed requests) of the information retrieval system;

i) fitness for execution of requests which can appear in the future;

j) standardization of technical means of information retrieval systems, etc.

Let us note that in different information retrieval problems one of the above-indicated characteristics can be taken as a criterion of optimization of work of the system, and others as limitations.

At any stage of functioning of an information retrieval system it is desirable to use a contemporary EVTsM which permits accelerating the time for performance of a requisition, increasing the accuracy and volume of obtained information, etc. During a more complete (exact) account of the character of functioning of the object.

Stages examined here of functioning of information retrieval systems and accordingly the structures of information retrieval systems are fairly general for a broad class of operational and designed information retrieval systems.

A SYSTEM OF AUTOMATIC DIFFERENTIATION OF DISTRIBUTION OF INFORMATION (SARI-1) ON CONSTANT INQUIRIES DEVELOPED IN TsBNTI AKIAE[1]

A. I. Nadtochiy, V. F. Kalinin, N. N. Mikheyev,
V. A. Voronin, V. I. Gostev, N. S. Denisenko,
and G. S. Chuba

Introduction

Contemporary rates of scientific and technical progress are connected with continuous growtn of flow of information materials and sharply increasing needs in information.

Needs in corresponding information for scientist and engineer appear on all stages of solution of a scientific-research or engineering problem.

It becomes evident that at the contemporary stage of scientific and technical progress providing of information is one of the decisive factors (and in many cases a bottleneck) to successful movement of scientific-research and research-design works.

In these conditions TsBNTI was faced with the problem of creating system of information service which would satisfy the need for concrete consumers, both with respect to time and with respect to subjects.

[1] Expansions unknown.

Solution of this problem began with creation of a reference and information fund the main component part of which would be unpublished materials - reports on scientific-research and research-design works conducted in foreign and domestic scientific research and design organizations. The principle of laying in a fund of these materials was accepted in connection with the fact that these materials present the greatest scientific interest, and they are not reflected in periodical publications.

With creation and accumulation of a reference and information fund there was taken on constant information service of a number of subscribers (scientific-research and design organizations, and also big scientific organizations) by which with the processing of a reference and information fund there had to be issued information on those thematic problems on which they are working at a given time.

To solve the problem there was developed a system of automatic differentiation distribution of information (SARI-1) with respect to constant inquiries with the help of the "Minsk-22" on the basis of descriptor language. The report gives the essence of this system.

I. Descriptor Dictionary (Thesaurus)

The list of descriptors accepted at the TsBNTI for the AIPS has the form of an alphabetic dictionary with indication of digital codes.

Digital codes were appropriated to simple numeration of elements of different lists entering the dictionary:

basic dictionary	0001-1231
supplements to it	1500-2000
glossary	6000-8000
supplements to it	4000-5000
countries	5000-5200
reactors	5200-6000
mass numbers	9000- ..

Thus, all the elements of the dictionary have four-digit decimal code, but are recorded in paired combinations in eight-digit code. Digital coding, instead of Euratom alphabetic coding is required for purely technical reasons. A descriptor dictionary is built on the basis of a thesaurus already proved with respect to frequency and a glossary accepted by Euratom and now recommended as an international tool within the bounds of MAGATE. In the course of practical application this dictionary in 1964-1965 there appeared the extreme necessity of its thematic and lexical expansion for a more complete scope of thematic interests of the proposed subscriber network of TsBNTI. For this it was necessary to supplement the Euratom thesaurus with additional descriptors from the dictionary of the Gmelinsk Institute (FRG) and the subject indicator of the American abstract journal Nuclear Science Abstracts (NSA). The volume of the basic dictionary increased from 1231 to approximately 1600 term-ideas.

Simultaneously from the Euratom thesaurus there were completely excluded two divisions of compound descriptors of the type

a) ferric oxide	iodine isotopes	sodium sulfates
uranium oxide	carbon isotopes	potassium sulfates
etc.		

with respect to 18 repeated descriptors for all chemical elements;

b) uranium-233	strontium-85	iodine-131
uranium-235	strontium-89	iodine-140
uranium-238	strontium-90	iodine-141
etc.		

for concrete isotopes of all chemcial elements. Instead of them in the instruction to the descriptor dictionary there is given a rule of free combination of any names of chemcial compounds with their concretizing chemical elements.

Oxide; iron	= ferric oxide
isotopes; iodine	= isotopes of iodine
chlorides; sodium	= sodium chloride, etc.

This made indexing more flexible and permitted composing such complicated descriptors as "uranium silicide," "molybdates of ammonium" and others, which was not in the Euratom thesaurus.

For coding isotopes with their mass numbers there was separated the 9th division of the decimal system. Mass numbers of any chemical elements are expressed by code with a 9 in front:

uranium = 235 has the code 1168.9235
strontium = 90 has the code 1070.9090, etc.

The terms of the Euratom glossary were almost completey introduced into our dictionary with all references provided for by Euratom. It also gave digital codes.

Furthermore, the dictionary includes coded lists:

a) countries, oceans and seas;

b) the most important atomic reactors;

c) the best known thermonuclear installations;

d) a small number of names of adjectives allowing concretizing such ideas as

temperature, low
temperature, high
uranium, natural
interactions, weak, etc.

As a result of all these reconstructions there was obtained a dictionary allowing indexing deeply enough any sources of information on very wide subjects and coding already prepared descriptor patterns (Euratom descriptions) and subject indicators (NSA).

The dictionary including glossary elements contains about 4000 term-ideas each of which can enter a paired combination if it has meaning and turns out to be necessary for reflection of fund material. The term-ideas "reactor" with code 0937 and "energy" with code 0869 give in our case the compound idea "power reactor" with code 0937.0869. The "analysis" and "chemistry" permit composing terms

chemical analysis with code 0056.1533 and
analytical chemistry with code 1533.0056

by a single rule: in the first place "that," and in the second "which."

In spite of supplements to the Euratom dictionary and reconstruction to adjust it to our concrete needs, the code system of the TsBNTI descriptor dictionary has a drawback: hierarchical or associative connection in the code reflection of descriptors is absent. For example, the ideas

nuclear reactor	0937
nuclear fuel	0459
active zone	0934
loop of heat-transfer agent	0259
retarders	0705
heat-transfer agents	0260

are not interconnected with respect to code, although they form a "field" of terms directly related to the idea "reactor" hierarchically common to them. The absence of classification connections in codes characteristic for all official codes in contrast to functional codes hampers information retrieval and makes necessary multiple probing of the fund with respect to the set of related descriptors, instead of one-two inquiry descriptors having a broader meaning.

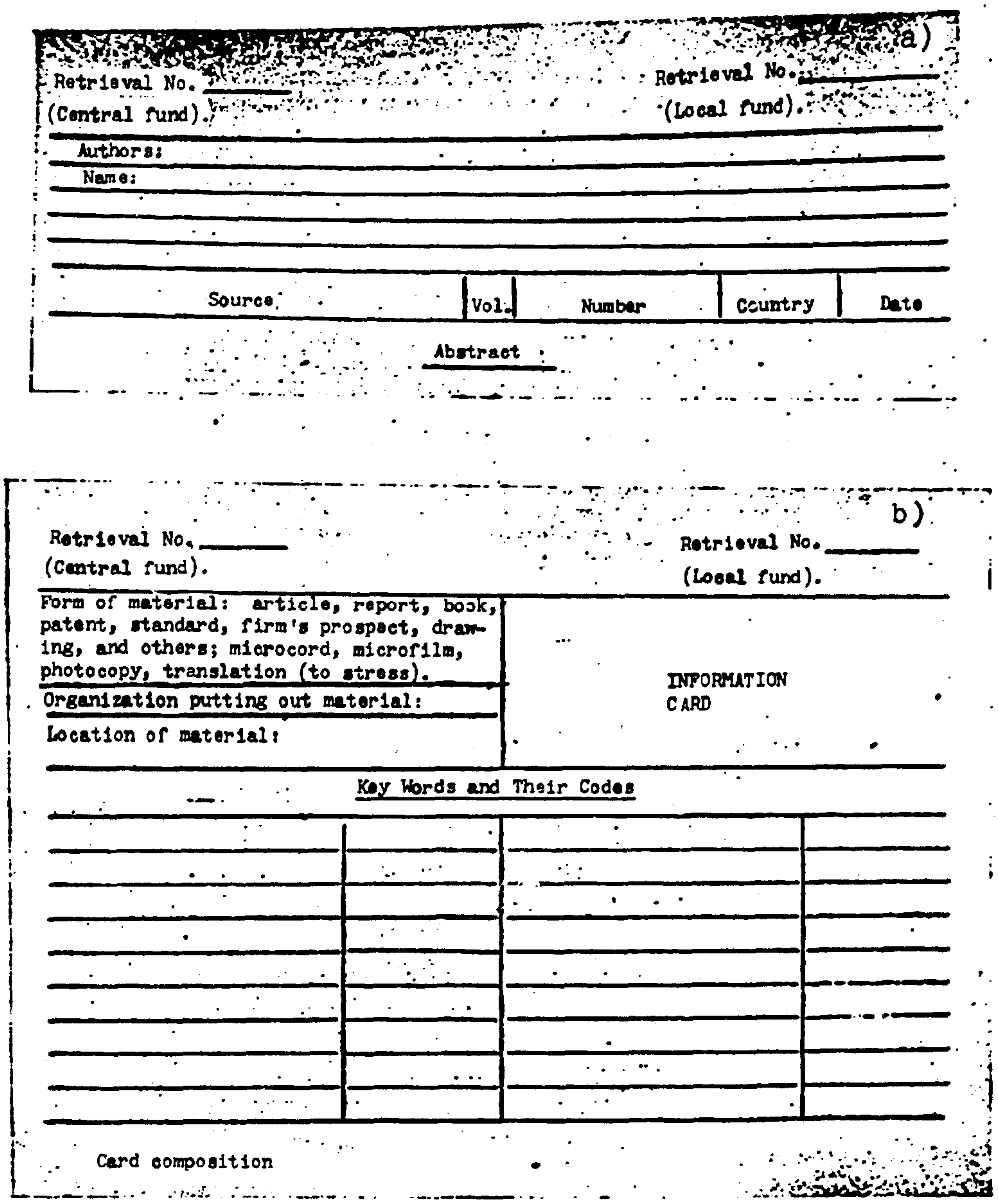
II. Processing Information Documents

A primary information unit which requires separate consideration and individual treatment is documents or their parts (a book chapter separate work in collection and so forth), treating of a thematically isolated subject. For example, in the report of the research center there are described various works in the field of metallurgy of uranium and other fuel materials. Each of the subjects touched is considered an independent document and processed individually.

An information retrieval cord with bibliographic data, annotation or abstract (if necessary and practically possible) and the descriptor pattern of the primary document (or an information unit of it) is a secondary document (VD) with a constant retrieval number and index of universal address of the document – the address assigned to the document during composition. For reports this is a literal abbreviation, the name of laboratory and the reference number of the report of this laboratory; for journalistic articles it is a four-letter code of the name of the journal on CODEN and numbers indicating number of volume, page and year. The retrieval number appears at all stages of processing, retrieval, and reproduction of primary documents. Contents of primary documents (or their units) are reflected in secondary documents in the form of a so-called descriptor pattern – a list of elementary or compound descriptors completely enough expressing the thematic essence of the given information unit. In a descriptor pattern there should not be nothing superfluous or secondary from the point of view of successful retrieval of a given document on inquiry. On the other hand, the pattern should contain all descriptors corresponding to the most important aspects of the present work having retrieval value.

In the accepted TsBNTI map on the face in certain places there are indicated exploration number of document, its authors and name in authentic writing, the translation of the name into Russian, annotation or abstract, initial bibliographical information and indication of the address of the abstract in NSA. On the other side of the card there is given information about the publishing character of the document, its form and fulfillment, place of publication and place of storage, and a list of necessary descriptors is given. The card is signed by its compiler. An approximate information card on a document in our fund is shown in Fig. 1.

The composed map is edited by a more experienced indexer, especially the translation of the name of the work and its descriptor pattern. Practice showed that even with the most conscientious relationship to indexing and translation the editor almost always



a)

Retrieval No. ____ (Central fund).

Retrieval No. ____ (Local fund).

Authors:

Name:

Source | Vol. | Number | Country | Date

Abstract

b)

Retrieval No. ____ (Central fund).

Retrieval No. ____ (Local fund).

Form of material: article, report, book, patent, standard, firm's prospect, drawing, and others; microcord, microfilm, photocopy, translation (to stress).

Organization putting out material:

Location of material:

INFORMATION CARD

Key Words and Their Codes

Card composition

Fig. 1. Form of information card. a) front; b) back.

improves the translation or list of descriptors. For editing there is a worker with a good grasp of the language of the original and technical indexing.

In the preface to the Russian-English variant of the descriptor dictionary there are contained methodical indications how to use the dictionary during indexing of the primary document. The result of indexing is the retrieval pattern of the original, i.e., the secondary document. The rules of indexing documents consist in sequential fulfillment of a number of operations.

1. Professional analysis:

a) analysis of all forms of presentation of the primary document from the point of view of their pithiness – title, abstract, thematic table of contents (divisions), complete text, and others,

b) check of the possibility of breaking up the document into independent information units (not to mix with rubricational division in subject cataloguing),

c) thorough study of contents of obtained information unit,

d) understanding of the system of basic ideas reflecting the contents of the separated part of the document, with use of the method of logical contrast within the limits of paired structure: "subject – about subject" (or "subject – subject"), for example: "power reactors" and "exploitation").[1]

2. Indexing:

a) composition of list of separated idea-terms. One should remember that use of binary terms (consisting of determination and determined) promotes high concreteness of expression; one should not use terminological compositions like "rim of wheel: or "American reactors," and it is necessary to have recourse to terminological unities, for example, "boiling-water reactors,"

b) finding in the descriptor dictionary lexical units for expression of the most exact equivalents of arranged ideas and

[1] The subject will be most frequently expressed in general-technical terms of the type "measurement," "breakdown," "production," "preparation," "treatment," etc.

"translation" or tracing of ideas on descriptors (in the form "simple descriptors") or descriptor syntagmas (i.e., "compound descriptors") using rules of grammatical synthesizing of a paired determining connection, for example, "boiling-water reactors" = "reactors boiling."

3. Coding:

a) coding of descriptors; two-four-digit digital code is used, i.e., a four-digit for the first component of the descriptor syntagma, and another four-digit code for the second component. In the absence of an attributive component, instead of the second component there is placed four-digit code, consisting of four zeroes (so-called "zero syntagma"). Between both components of two-four-digit code there is placed a sign of coupling, designated by a point.

As can be seen from the rules, in retrieval language serving for recording the contents of a document, there are used now only two forms of coupling: – logical (or "semotactical") – between syntagmas, i.e., expressed by opposition "subject-theme"; – syntactical – between components of a paired descriptor syntagma – i.e., coupling "determined – determination" (so-called "postpostitive attributive syntagma").

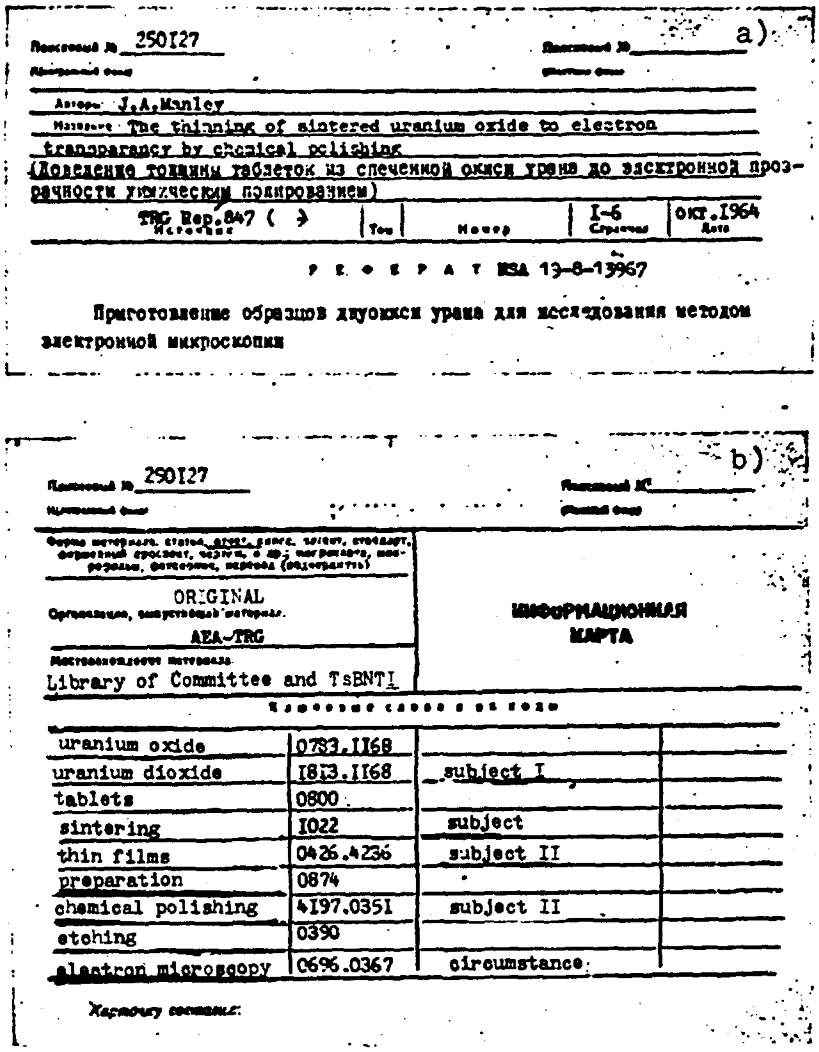
Inside the actual pattern separate descriptors are disposed in groups in accordance with basic elements; subject, theme, and circumstance (or condition). This it is necessary so that a similar "telegraph" recording in some measure replaces or duplicates the information abstract. An example is the pattern depicted in such a way on applied map No. 250127 in Fig. 2.

III. Processing of Inquiries

1. Form of Inquiry

In the actual beginning of creation of SARI-1 in TsBNTI there were determined consumers[1] of information, selected in the process

[1] In TsBNTI the consumer of information which is whole by an independent establishment is named "object SARI" with its registration number.



a)

Поисковый № 250I27 Поисковый №

Инвентарный фонд

Авторы J.A.Manley

Название The thinning of sintered uranium oxide to electron transparancy by chemical polishing

(Доведение толщины таблеток из спеченной окиси урана до электронной прозрачности химическим полированием)

TRG Rep.847 ( )			I-6	окт.I964
Источник	Том	Номер	Страницы	Дата

Р Е Ф Е Р А Т NSA 19-8-13967

Приготовление образцов двуокси урана для исследования методом электронной микроскопии

b)

Поисковый № 290I27 Поисковый №

Инвентарный фонд

Форма материала: статья, [illegible] (подчеркнуть)

ORIGINAL

Организация, выпустившая материал

AEA-TRG

Местонахождение материала

Library of Committee and TsBNTI

ИНФОРМАЦИОННАЯ КАРТА

Ключевые слова и их коды

uranium oxide	0783.II68		
uranium dioxide	I813.II68	subject I	
tablets	0800		
sintering	I022	subject	
thin films	0426.4236	subject II	
preparation	0874		
chemical polishing	4I97.035I	subject II	
etching	0390		
electron microscopy	0696.0367	circumstance	

Характер сообщения:

Fig. 2. Form of filled infoımation card. a) front; b) back.

of work of the AIPS [Automatic IPS] — scientific research and planning and design organizations of the State committee.

Every consumer was assigned a conditional number (from 01 to N...) and there was sent notification about acceptance of a given enterprise in the number of subscribers SARI-1, instruction according to formulation and descriptor description (or descriptor scanning) interesting this enterprise of subject. The instruction offered the following form of inquiry sent to the TsBNTI.

The usual card, the dimensions of which were indicated beforehand, had to be filled from two sides. One one side there was supposed to be written a so-called "formulation of inquiry," i.e., the name of the interesting subject in natural (Russian) language, and on the back the "descriptor description of the inquiry." For the latter there was used an applied descriptor dictionary. Assignment of a "descriptor description," which is more correctly called "descriptor scanning," consisted in that the "description," first, brought to a group of TsBNTI inquiries a more detailed expression of a subject, and secondly, played the role of a normalized AIPS language. The latter is very important since this was normalization not only in the direction of retrieval language, but also normalization in the terminological sense.

Experiment showed that descriptor description of inquiries by subscribers turned out to be low-quality due to inability to use descriptor language and complexity of the actual method of indicative mood reviewing of subject, which in point of fact is "descriptor description." For example, we will take one of the inquiries sent:

> Formulation of inquiry: "Obtaining by way of gas electrolysis of volatie compounds of refractory metals."
>
> Descriptor description: "Refractory materials."

Here descriptor description does not definitize the subject of inquiry but expands it, not even mentioning that "refractory materials" are not only metals; the class of refractory metals includes all metals with a melting point higher than the melting point of iron. In particular, during the analysis of this inquiry in TsBNTI there were 16 descriptors in the list: molybdenum, tungsten, niobium, tantalum, titanium, vanadium, chromium, zirconium, ruthenium, rhodium, palladium, hafnium, rhenium, osminum, iridium, platinum.

Later there was accepted the decision to establish feedback with subscribers of inquiries. The most effective method of such contact is direct conversation with the interrogator. Started in 1966, this practice has not yet been brought to completion but has already given good results. In particular, after conversation with the subscriber of the above-cited, out of sixteen refractory metals there remained only four, which considerably decreased input volume during retrieval and, consequently, lowered the cost of machine work.

Distribution of information documents by constant inquiries is carried out periodically with the putting into the AIPS of new information material immediately after obtaining the form of the inquiry. The form of inquiry obtained by TsBNTI in book of constant demands and from this moment becomes as the "information inquiry questionnaire" of the given subscriber. In TsBNTI retrieval by constant inquiry is regular, and, since the reference and information fund (SIF) SARI grows with the processing of entering materials by a comparatively small number of informant-indexers, single input of documents into the AIPS cannot be great. At present each AIPS gets 300 documents in a week. The retrieval system is fed 182 constant thematic inquiries from a number of organizations.

2. Study of Subjects of Inquiry

In connection with the fact that it is necessary clearly to concretize the expression of the inquiry, the subjects of the latter turns out to be almost the most important aspect of work during the analysis of entering inquiries. During descriptor treatment of the inquiry before us, there came up not only the question of *thematic referredness* but also the problem of *expression* of these contents, where the latter is immeasurably more important, since during retrieval by the informant there stands the question *not to reproduce a phenomenon* physically and even *not to explain* its meaning to any person but to *express* it in the language of descriptors. "Descriptor description," sent by the author of the inquiry in this case only insignificantly helps during the analysis of the subject of inquiry since it most frequently does not fulfill its assignment. Even moreover: if "formulation of inquiry," being surface, often leads to error, then "descriptor description" gives too many variants in an attempt to relate the inquiry to a branch.

Inasmuch as the subscriber must give a clear account of the subject interesting him, foresight of usual difficulties during "formulation" and "descriptor description" determined the necessity during selection of descriptors of access to the thematic divisions of the dictionary. In connection with this it is possible to give two case, which turned out to be the most typical.

A. "Formulation" of Inquiry has "Tight" Filling

Example: "Obtaining of different classes of organic compounds containing radioactive isotopes. Questions of introduction of radioactive isotopes in proteins, amino acid, nukleodizy [no translation found], nucleotides, nucleic acids, vitamins, hormones, antibodies, benzene, naphthalene, and their derivatives."

Descriptor description
Methods of obtaining
Organic compound
Radioactive isotopes

B. Formulation" of Inquiry has "Weak" Filling

Example: "Zirconium and its alloys."

Descriptor description
Zirconium
Alloys of zirconium
Corrosion
Oxidation
Oxides
Metallurgy
Diagram of states
Intermetallic compounds
Diffusion
Mechanical properties
Adsorption
Hydrogen
Soldering
Welding
Heat treatment
Irradiation

The difference between both examples is absolutely evident, but regarding thematic referredness, in the first case it is easier to trace with respect to "formulation of inquiry," and in the second with respect to "descriptor description." With such an approach to thematic analysis for us the subject of the theme is one of the ideas selected from the formulation of the inquiry, and the theme is the fact that is expressed by one of ideas contained in the "descriptor description" and designating the process of introduction of radioactive isotopes into an organic compound.

It is not difficult to note that case "B" requires absolutely the reverse approach. Here the theme is the general content of the "descriptor description," i.e., "metallurgy," and the subject is "zirconium" or "alloys of zirconium" (or their synonyms).

From the point of view of retrieval language every inquiry is considered an expression which must contain a subject and what is said about this subject, i.e., its theme. "Representatives" of a subject are usually selected from a number of concrete, subject ideas, but "representatives" of a theme are expressed by names of processes, operations, phenomena, states, or their sets (in example "B" such a "joint" idea is descriptor "metallurgy," which causes determination of it as themes). In connection with this the knowledge of a logical system of ideas of a dictionary of retrieval language is absolutely necessary, or it must at least be grasped intuitively, if the researcher has no skills of formal analysis of language. With respect to the two examples given with "tight" and "weak" filling one should say the following: during the shaping of an inquiry extremes on neither side are permissible. An inquiry with "tight filling of formulation" is no longer an inquiry, but many inquiries with matching themes and, consequently is subject to breaking up (case A). An inquiry with "weak filling of formulation" (case B) is an inquiry containing the subject of an expression, but without a theme, which must be looked for in "descriptor description." Therefore, the following rules cover composition of an inquiry:

1) construct the inquiry in the form of an expression containing one subject and one theme in natural (not descriptor) language;

2) in the descriptor description under the heading "П" enumerate all descriptors having to do with the subject of expression. Select

descriptors from dictionary, but new ideas may be introduced if necessary;

3) the same as in the above paragraph, for the theme headed "T."

Only if these rules are observed, will an inquiry have so-called "transparent structure." In another form inquiries sent to the TsBNTI, did not respond to this formula but made it necessary to conduct "structural analysis of the inquiry."

3. Structural Analysis of an Inquiry

Structural analysis emanates from the preceding procedure of thematic analysis and has as its purpose to turn the initial inquiry into a group of its modification-subinquired.

The structure of an inquiry is expressed in the form of a relation identical to logic function "and," i.e., assumes "single-placeness" or "simultaneity" of both parts of the basic structure, i.e.,

$$\Pi \wedge T.$$

where sign Λ transmits the "and" relation. Without going into a detailed account of this question, let us note only that structural analysis involves two forms of actions or operations.

The first operation is <u>horizontal scanning of the structural formula</u>, i.e., calculation of all components of both the subject and the theme of an inquiry. Thus, for example, for case "B," given in the preceding paragraph, scanning will have the following form:

П	Λ	T
zirconium, zirconium alloys		metallurgy, corrosion, diffusion, oxidation, mechanical properties, etc.*

*For the sake of neither horizontal nor vertical scanning is carried to completion.

The second operation is "vertical scanning." This means that from the thesaurus there have to be selected all synonyms of both the subject and the theme. For example:

П	∧	Т
zirconium (zirconium alloys)		metallurgy (corrosion) of oxide
1. zircalloy		1. heat treatment - 1. of dioxide
2. alloy A-816		2. cold treatment
3.		3. pressing
4.		4.
.		.
.		.

As a result of horizontal and vertical scanning we have two subsets — subset "П" and subset "T," and within the limits of each of them there acts the relation of "various-site," expressed by the "or" function.

In the course of structural analysis there can be clarified false associations (i.e., combinations not confirmed by the meaning of the inquiry or the common meaning). This makes it possible, first, to avoid uninformative inquiries, and secondly, to articulate "complicated inquiries" into subinquiries.

4. Modification of Inquiry and "Working Inquiry"

Modification of an inquiry is its breakdown into variants with respect to certain considerations determined by its structure or the possibilities of descriptor presentation. Inasmuch as the subject or theme can be represented by totalities with internal "or" relations, the number of possible combinations ПΛТ can be very great. Thus, for example, if there are 10 members, theoretically it is possible to take 10 × 10 = 100 modifications (in practice fewer of them). An inquiry entering the number of real modifications introduced into the AIPS, is called a working inquiry[1] in contrast to the initial thematic inquiry.

[1] That is, the inquiry in that form in which it is introduced into the retrieval system.

Indexing of inquiry. Indexing of inquiries is the next stage of treatment and at the same time the last conditon determining quality of SARI work and AIPS effectiveness.

Indexing can be conditionally called "marking of the inquired subject" and provision of the inquired theme with criteria according to the document in the file will be identified during retrieval. Hence, naturally, descriptor presentation of an inquiry must meet certain requirements, nonobservance of which will lead to information losses or information noise at the output of automated retrieval. Noise can be uncovered by examining a set of issued documents and comparing it with the inquiry, but this is considerably easier than examining the whole file.

We are not trying to get rid of noise altogether but to minimize it. The fight for IPS quality is a fight for lowering of noise, inasmuch as losses need not come under discussion at all since there should not be any of them or at worst they have to be insignificant. For struggle with losses in the order of the document list there are several procedure which, in the end, lead to attracting paradigmatic connections of the dictionary, i.e., analyzing connections between descriptors.

The basic rule of control of search is the more identification criteria (descriptors) in an inquiry the less probability of their completely matching the descriptors of the retrieved document.

The number of descriptors (i.e., the depth of retrieval) is a very important AIPS parameter since increase in the depth of retrieval, i.e., increase in the accuracy of the inquiry can lead to losses; decrease in depth decreases losses but can increase noise. Such, unfortunately, is one of the regularities of probabilistic retrieval. Usually there is found (after a certain number of experiments) an average depth of inquiry. At the TsBNTI for SARI-1 inquiries the average depth is 2-3 descriptor (connected by an "and" function for quantitative calculations — 2.5) if the average number of descriptors in the document is 15. Thus, full coincidence of an inquiry with a document is *most probable* when the number of descriptors of the inquiry is to the number of descriptors of the document as 1 is to 6.

The hard part of indexing is reaching the average parameter of the inquiry *without distorting the contents of the inquiry*. The fact is that only *in structural relation* does an inquiry consist of two members — Π and T. But a member of the syntagma (syntactical combination) of an inquiry in most cases does not match an individual *unit* of the descriptor dictionary. More frequently either member Π and T are expressed by two units, but if an inquiry consists of four descriptors. then there appears a threat of losses. For example:

Inquiry: "Economics of production of uranium fuel"
Structure: Π — (uranium, fuel) as Π_1, Π_2
T — (economics, production) as T_1, T_2
Indexing: Uranium — Fuel — Production — Economics

As can be seen from the example, the working inquiry consists of four descriptors, which threatens information losses since there surely will be found a document in which information is contained on this subject, and the descriptor "Production" or the descriptor "Economics" is absent. Therefore, the inquiry in the form obtained after analysis is *modified* into two synonymic constructions:

a) uranium — fuel — production
b) uranium — fuel — economics

Instead of one thematic inquiry there were obtained two sub-*inquiries*.

Another example is the still more complicated case in which inquiry density is so high that losses are inevitable:

Inquiry: "Fast-neutron power reactors with high burn up of fuel and increased reproduction."

Here the author of the inquiry does not explain just what about this reactor he needs, and, therefore, we consider that he needs everything on this class of reactors, and consequently:

Structure: Π_1 – power reactor
Π_2 – fast neutrons
Π_3 – high burn-up of fuel
Π_4 – increased reproduction
T – 0
Indexing: Power reactor – fast neurons – burning
out – reproduction

When only one member of the syntagma is expressed by more than two descriptors, it is possible to say beforehand that the probability of finding the document is insignificant. In this case we furthermore arrive at the modifications

M_1 – power reactor – fast neutrons – burning out
M_2 – power reactor – burning out – reproduction
M_3 – power reactor – fast neutrons – reproduction

Besides modification we usually use the so-called "method of substitution" or replacement. This is possible only if descriptor scanning of the inquiry after indexing allows according to the laws of descriptive lexicology replacing the "sum of ideas" with one idea in the thesaurus. Here instead of four descriptors of scanning at the same value (with distortion of meaning, of course) substitution (S) is possible:

S_1 – breeders

Substitution is expanded here basically to "reactors" and "increased reproduction of fuel," i.e., substitution is partial, but in this sense the purpose of introducing variants, subinquiries, modifications, etc., is to collect all partially synonymic expressions consisting of not more than three descriptors, so that their sum completely covers the meaning expressed in the inquiry. Each of the variants can give and surely will give noise, but then losses are avoided, which inevitably appear if the machine is fed complete scanning of the inquiry. Incidentally it is necessary to note that such cases as the last example are few. It remains only to add that substitution is only an intermediate procedure in the sense that the substitute ("replacement") can be combined with other "unreplaced" scanning descriptors such as, for example:

M – breeders – fast neutrons

However, the number of modifications is increased to such dimensions that it is necessary to remember to save the volume of memory of input to the ETsVM.

Correction of inquiry. Everything said above brings us to the final stage of processing inquiries – correction of working inquiries with respect to analysis of feedback. Inasmuch as AIPS effectiveness is estimated as a percentage of noise information and information losses, output (delivery of documents) AIPS gives to us that material which contains data for corrections of inquiries on any of the enumerated stages of treatment. It is necessary to note that error in the inquiry or its inaccuracy on the most first stage (presentation of the inquiry by the subscriber) does not depend on the work of the group of TsBNTI inquiries and they are removed only during conversation with the subscriber during personal contact. Obviously, taking measures to remove error of this type is very important, since if error is allowed in the very beginning of the process of work, the error will spread with passage through subsequent phases of treatment, where its dimensions will be greater the more intermediate operations there are up to the moment of input.

Regarding the correction of inquiries on remaining phases of treatment, we do this regularly with the accumulation of material by way of study of output tabulograms and comparison of output data with the retrieval array and inquiries in a given stage of treatment.

5. Evaluation of Informative "Weight"

After the above it is easy to see that formulation of an inquiry is subordinated to the language system on the one hand and the system of descriptor retrieval on the other. This means that all operations accomplished with inquiries are in point of fact one of the most important methods of controlling the retrieval process. But to keep down noises and losses in TsBNTI retrieval subsequently there is to be used one of the procedures — "evaluation of informative weight" — of the descriptors of the inquiry which consists in the following.

Ideas introduced into the group of inquiry descriptors possess various degree of information recognizedness. Of all the descriptors some have very direct linage with the inquiry (subject or theme), others have mediated linkage, and a third group is so weak what a descriptor can be considered optional (i.e., both its presence and its absence in the inquiry is grasped as controversial). Now, if we dispose all inquiry descriptors by method of "diminishing criterion" (the so-called "gradual series"), then in the first place we have the most "informative" descriptor and in the last the least informative, for example:

Inquiry: Investigation of the contents of strontium-90 in the atmosphere.

Descriptor description: 1. strontium-90
2. atmosphere
3. contents
4. investigation[1]

Gradual series: strontium-90 — atmosphere — contents — investigation

The series contains four components. If we assume that the number of gradations must not exceed four, as in this case, then we could "evaluate" the informativeness of every descriptor by its place in the series, designating thereby its informative valence or so-called "informative weight" (U).

1. strontium-90, U-4
2. atmosphere, U-3

[1] For simplicity of account we take the "ideal" descriptor description, omitting analysis, expounded in the preceding subdivisions.

3. contents, U-2
4. investigation, U-1

The total sum of weight is 10 units of valence. Taking this number as the initial constant, we can assume that correct delivery of information is within limits from 7 to 10 (or in other limits which are set by a series of checks of statistical order). It is not difficult to note that, setting the optimum number at 9, we manage only without descriptor "investigation," which in fact insignificantly affects the outcome of retrieval. Thus, we will obtain one more method of controlling retrieval.

IV. The Structural-Functional Diagram of the SARI-1

The SARI-1 AIPS is a descriptor retrieval system working on differentiational distribution of information. This means that a set of inquiries and the basic principle of the work process consists in series comparison of "descriptor descriptions" of documents with "descriptor presentations" of inquiries. The result in the form of a tabulogram indicates that a given inquiry corresponds to a given subset of documents. A tabulogram is a list of digital codes of a finite set of inquiries which registers the subset of the retrieval (i.e., identification) numbers[1] of those documents the descriptor description of which completely includes the descriptor presentation of the given inquiry.

In the expounded meaning SARI-1, as an AIPS, belongs to a class of retrieval systems, i.e., during work accessing an array of documents consists in two phases of the process:

— comparison with a finite set of inquiries and

— differentiational distribution in linkage.

[1] Universal addresses in general.

An AIPS is a basic functional organ of the SARI-1 and occupies in it a place, which can be characterized as a "subsystem in a system," although the property of these two systems by far not are identical. As shown above, SARI is differentiational distribution of information in which it was decided to use the "Minsk-22" as technical means playing the role of an automatic retrieval device. Automation of such distribution was recognized as expedient because at the TsBNTI there was set the problem of processing a large information fund. Let us enumerate basic functioning parts of SARI.

The flow of information in the form of reports, articles, abstracts, etc., is sent to the library of the information fund.

Processing of information (primary document "PD") for the purpose of presenting it in a form convenient for retrieval from the point of view of AIPS requirements (secondary document "VD") is carried out in the division of the reference and information fund by an indexing group in which all necessary data are introduced into the card in the form of formal criteria – indices (retrieval numbers, addresses, descriptors with digital codes, and others).

Form VD is an information card (form No. 1). Removal of data from the secondary document (indices) to the retrieval sample of the document (POD), for example to a punched card, punched tape, or magnetic tape (depending upon input method), is carried out in the punching group of the section of mechanization of information processes. Reception, analysis, processing, and modification of inquiries and their conversion into VD form are carried out in the group of inquiries of the section of the reference and information fund.

The SARI structure also includes:

A network of SARI objects – subscribers of information service. On the first stage we were limited to a comparatively small number of subscribers – the leading scientific and engineering-technical workers of scientific research institutes and design bureaus of the basic thematic directions of our branch of science and technology. The number of such thematic inquiries (or "inquiry-subscribers") obtained at the TsBNTI was 182.

A network of peripheral information services directing information documents (PD and or VD) to the central fund. In the first place it was decided to deposit in the fund reports of foreign centers and firms working in the field of atomic science and technology, and also open reports of scientific research institutes and design bureaus of the Glavatom [Main Administration for the Use of Atomic Energy] system.

Solution of problems of automated information retrieval (AIPS) – development of algorithm and program and practical fulfillment of retrieval – is carried out by the division of mechanization and automation of information processes (OMAIP) of the TsBNTI and the computer center of the Central Statistical Administration of the RSFSR, where we are renting a "Minsk-22."

The processing of output tabulograms, which are "output production" of AIPS, at present is carried out by a special "group of processing of output" of AIPS.

Feedback. A copy of the information card found according to the inquiry is sent to the subscriber with a special breakaway coupon ("stub") containing a mark characterizing the reaction of the buyer for the document in the fund, for example: "The document sent interests him a) completely, b) partially, c) information on the given question is obtained for the first time, d) material is known" or "He needs a) abstract, full material; b) original microfilm or microcard" or "The material obtained does not correspond to the inquiry." The feedback coupon is sent back to the TsBNTI where part of them – with positive answer – goes to the group of distribution, and part – with negative answer – enters the group of inquiries for analysis of unsatisfactory output and correction of input.

Such is the structural-functional diagram (Fig. 3). As can be seen from the diagram, the function of separate sections coincide not by form or assignment of material (report or information card,

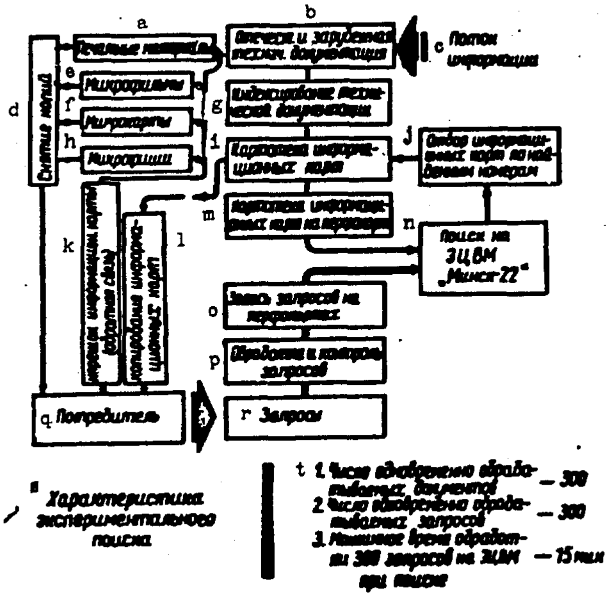


Fig. 3. Diagram of SARI-1 work.
KEY: (a) Printing material; (b) Domestic and fcreign technical documentation; (c) Flow of information; (d) Removal of copies; (e) Microfilms; (f) Microcards; (g) Indexing of technical documentation; (h) Illegible; (i) File of information cards; (j) Sampling of information card by found numbers; (k) Stub of informational card (feedback); (l) Copying of information cards; (m) File of punched information cards; (n) Retrieval on the "Minsk-22"; (o) Recording of inquiries on punch-cards; (p) Processing and checking of inquiries; (q) User; (r) Inquiries; (s) Characteristic of experimental retrieval; (t) 1. Number of simultaneously processed documents. 2. Number of simultaneously processed inquiries. 3. Machine time of processing 300 inquiries during retrieval.

as form VD; demand or information material, etc.), by its presentation by a method, and by methods of its processing.

In the beginning of this division we talked about "descriptor description of a document" and "descriptor presentation of an inquiry." The meaning of these two ideas is that information materials and inquiries can equally be considered two varieties of document, of which one is recorded in natural language, and the second of which is processed in digital codes for moving retrieval data to the retrieval sample of the document.

The secondary form of the inquiry somewhat differs from the secondary form of the information document. The difference is that while an inquiry is being processed it (as a rule) has no bibliographic and factographic data; therefore, the corresponding secondary form contains only an identification number and codes of descriptors expressing the theme inquired about (see the section "Processing of inquiries" in greater detail).

V. Machine Realization of SARI-1

1. Introduction into System. Form of Presentation of Primary Information

Documents presenting a volume of current information on the one hand, and constant inquiries reflecting the interest of users, on the other hand, are put, into the system.

The abstract of the document in the form of code-descriptors is recorded in an information card. Descriptors are described in eight- and four-digit digital codes, where the first main part of the eight-digit code appears as an independent code on the information card:

0746.0000
0746.9017

For input to the "Minsk-22" the contents of the information map (registration number and code-descriptors) is recorded on punched cards in the decimal system. Every punched card contains 9 descriptors, and the average number of descriptors of a document is 20, consequently, on every document 2 (two) punched cards are recorded. Both copies of punched cards of one document contain one registration number of the document. (The program provides for recording documents on punched tape when necessary.) The system is fed 455 punched cards or approximately 270-300 simultaneously.

Constant inquiries of users are broken up into subinquiries. Every inquiry includes not more than 40 subinquiries. An individual subinquiry contains 2-3 descriptors and is recorded for punch-card input. The machine is simultaneously fed a group of 300 subinquiries (punched cards). Formed and definitized groups of inquiries are recorded on magnetic tapes.

An auxiliary program provides for printing a punched array of punched cards of documents (inquiries) and the obtaining of the check sum on a given group of cards. Printed data permit restoring punched card in case of loss or damage.

Unit I of the main program provides:

1) input of documents to the "Minsk-22" (punched card, punched tape).

2) regulating of descriptors within every document,

3) creation of information tables about documents – length of retrieval pattern of every document and initial address of its location in MOZU [Magnetic working storage] (Form see Appendix).

Unit II provides:

1) input of inquiries to the machine (punched cards, magnetic tape),

2) regulating of descriptors within each inquiry,

3) creation of information tables about inquiries – length of retrieval instruction of every inquiry and initial address of its location in MOZU (Form see Appendix).

2. Storage of Information in Machine. Accumulation of Information

Provision is made for a direct form of storage of information in the machine: recording presents a sequence of descriptors located behind the document number

```
+580011479    +580011480
+000560000    +000790000
+000561533    +002690000
+001650000    +003020000
+001900000    +003020108
+006290000    +004760000
+006570000    +004890000
+008240000    +004899198
+008640000    +006900000
+009930000    +009140000
+010270000    +009140108
+012040000    +016520000
+012040993    +050530000
+015450000
+015840000
+060530000
```

Every number and code-descriptor goes with a 37-bit cell. They are stored in binary-decimal code. The difference between a code-descriptor and the code of a document number is that there is a 5800 or a minus sign (-) in the highest bit of the latter.

On operational storage of punched cards of documents for retrieval there are assigned 4096 cells (words) of unit II of MOZU. On operational storage of inquiries for retrieval there is assigned half of unit I of MOZU — 2000 words; documents and inquiries are located in order of input. The necessary input condition is an array of punched cards regulated by numbers of documents (inquiries). Provision is made for accumulation of information on punched cards.

3. Method of Retrieval

Use is made of a simplified algorithm based on simple comparison of descriptors of inquiry and document. By the table of information of inquiries (T.Z) there is selected the first subinquiry, and in sequence by the table of information of documents the retrieval instruction of the inquiry is crossed with the retrieval pattern of every document. The resultant list of crossing is compared with the inquiry and there is carried out echeloning of the delivery of the answer.

Form of delivery of answer:

1. If all (n) descriptors of an inquiry match all or some descriptors of a document, the number of the given document is issued

in the 1st echelon.

2. If n-1 descriptors of an inquiry match all or some descriptors of a document, the number of the document is issued in the 2nd echelon.

3. If n-2 descriptors of an inquiry match, the number of the document is issued in the 3rd echelon. (Program provides for cutoff of the 2nd and 3rd echelons through a key; thus during exploitation of the system, a 3rd echelon providing for the matching of n-2 descriptors of the inquiry with the descriptors of the document ceases to be necessary.)

4. Delivery of Information

Retrieval results are printed on an ATsPU (alphameric printer).

In the corresponding columns in the binary-decimal system there are printed:

a) the number of the inquiry and its retrieval pattern (descriptors);

b) numbers of documents by echelons of delivery and descriptors of documents which matched the descriptors of the retrieval pattern of the inquiry.

5. Reproduction and Distribution of Documents

As can be seen from the flow chart (Fig. 3), after the tabulogram has been processed, from the fund there are selected information cards corresponding to the inquiry. They are manually selected from the fund (cards are decomposed in order of retrieval numbers). Removed cards are examined by a specialist for evaluation of contents (to decrease noise).

Information cards selected for the user are sent to an "ERA" installation for reproduction. From the original-mock-up there are prepared on the "ERA" two forms (front and back), from which there are reproduced impressions of blanks.

After obtaining a copy of the information card with a feedback stub, the latter are distributed to information users. From the information user "stub-cards," "feedback" enter the group of inquiries where the inquiry is corrected – in case of a negative answer – materials are selected for reproduction. The order and originals are transmitted to the microfilming laboratory or to the section of operational reproduction in the form of works which must be conducted. Microfilming is done on a UDM-2.

Laboratory technology has been developed at the TsBNTI <u>for removal of copies from microcards and documents</u>.

Single copies are removed by means of electrography. Copies from small individual materials are taken by means of electrography on the "ERA" (for sheet materials) and the "Electrophot" for stitched materials.

During the time of field testing of the SARI-1 system during 1.5 years, about 2000 originals were reproduced from stubs.

VI. Determination of Effectiveness of Automated and Differentiational Distribution of Information (SARI-1)

We analyzed a comparatively small array of documents; obtained data permitted reaching a simple conclusion with respect to ways of improving the investigated system.

The basis of the given method of calculation is the method of A. V. Sokolov. According to this method parameters characterizing the accuracy of IPS work are loss of useful information L and information noise N%, which are determined by the corresponding formulas:

$$L = \left(1 - \frac{S}{S_{\Sigma}}\right)\cdot 100\%,$$

$$N\% = \frac{N}{p}\cdot 100\%,$$

where S_Σ is the number documents introduced into the IPS completely or partially satisfying a certain inquiry (containing useful information on a given inquiry); S is the number of documents issued by the IPS containing useful information (part from S^{a});. N is the number of documents issued by the IPS not containing useful information on a given inquiry (information noise); P is the total delivery of IPS on a given inquiry (S + N).

Thus, S/S_Σ characterizes completeness of delivery, and N accuracy of delivery. Completeness and accuracy of delivery are basic characteristics determining effectiveness of IPS work.

The values S, P and N in tables giving analysis of delivery of documents by the system (Tables 1-5) are represented in the form of fractions the numerators of which are the number of documents issued by the system in the 1st echelon, and the denominators of which are the number of documents issued by the system in the 2nd echelon. The criterion of delivery of the document in the 1st echelon is the presence in descriptor description of this document of all descriptors of the inquiry, in the absence of one inquiry descriptor in the descriptor description of the document this document is issued by the system in the 2nd echelon.

Table 1. Inquiry 12001. Power reactors with nuclear superheating of steam: construction, parameters, operational experience (energy reactors, power installations, overheating, nuclear overheating, superheaters) S_Σ = 8.

n/s	Descriptor description	S	P	N	Calculations of losses and noises
1.	power reactors + overheating	0/8	0/32	0/24	$L=0$; $N\%=75\%$
2.	power plant + overheating	0/7	0/11	0/4	$L=12,5\%$; $N\%=36.4\%$
3.	power reeotors + nuclear overheating	0/0	0/21	0/21	$L=100\%$; $N\%=100\%$
4.	plants + nuclear overheating	0/0	0/3	0/3	$L=100\%$; $N\%=100\%$
5.	power reactors + superheaters	0/8	0/8	0/0	$L=0$; $N\%=0$
6.	power plants + superheaters	0/8	0/8	0/0	$L=0$; $N\%=0$
	Altogether	0/8	0/32	0/24	$L=0$; $N\%=75\%$

Table 2. Inquiry 12011. Operational experience of boiling-water reactors (boiling-water reactors, exploitation, reactors, boiling) $S_{\Sigma} = 9$.

n/n	Descriptor description.	S	P	N	Calculations of losses and noises
1.	boiling-water reactors + + exploitation	0/1	0/16	0/15	L=85,7% N%=93,75%
2.	boiling + reactors + exploitation	0/1	0/16	0/15	L=85,7% N%=93,75%
		0/2	0/22	0/20	L=71,4% N%=90,9%

Table 3. Inquiry 12019. Zirconium and its alloys (zirconium, zircalloy, alloys of zirconium) $S_{\Sigma} = 12$.

n/n	Descriptor description	S	P	N	Calculations of losses and noises
1.	zirconium	4/0	14/0	10/1	L=66,7%; N%=71,4%
2.	alloys of zirconium	6/0	6/0	0/0	L=50%; N%=0;
3.	zircalloy	3/0	8/0	5/0	L=75%; N%=62,5%
		12/0	27/0	15/0	L=0; N%=55,6%

Table 4. Inquiry 12029. Diffusion of products of fission of uranium dioxide at high temperature (uranium dioxide – fission products – diffusion – high temperature) $S_{\Sigma} = 5$.

n/n	Descriptor description	S	P	N	Calculations of losses and noises
1.	dv. [unknown] uranium + fission products + diffusion + high temperature	1/2	1/2	0/0	L=40%, N=0
		1/2	1/2	0/0	L=40%, N=0

Table 5. Inquiry 25005. Processing and burial liquid waste at an AES [Atomic Electric Power Plant] and enterprises (liquid waste, removal of waste, waste treatment) $S_{\Sigma} = 20$.

n/n	Descriptor description	S	P	N	Calculations of losses and noise
1.	liquid waste + removal of waste	0/17	0/19	0/2	L=15%, N=10,5%
2.	liquid waste + waste treatment	0/13	0/14	0/1	L=35%, N=7,14%
		0/19	0/21	0/2	L=5%, N=9,5%

On the given inquiry the system has no losses and 24 documents not satisfying it are issued due to the presence in their descriptor description of the descriptor "power reactors."

Twenty documents not satisfying the inquiry are issued by the system on the descriptor "exploitation" on the 1st modification and the descriptors "exploitation" and "reactors" on the 2nd modification.

Of nine relevant documents only two are issued, and retrieval numbers 06209, 06680, 07610, 09097, 09098, 09219, and 09223 are lost.

Direct comparison of the descriptor descriptions of the inquiry and documents mentioned above, permits uncovering the cause of loss of these documents by the system.

In all the above-mentioned documents in the descriptor description there is shown the descriptor "boiling water" and separately "reactor." In the descriptor description of the inquiry and its modifications there are shown the descriptors "boiling-water reactor" and "boiling," "reactor."

The same fact or phenomenon is described by indexers of inquiries with one bunch of descriptors, and by indexers of documents by others. Therefore, one should introduce an understanding with respect to application in descriptor descriptions of the document and inquiry of the synonymic construction, most completely responding to the meaning; this understanding should be reflected in the instruction with respect to indexing of documents and inquiries.

The descriptor description of inquiry 12019 assumes delivery of material by the formula "all on zirconium and its alloys." If issued documents are examined from this point of view, then all 27 documents should be considered relevant, and information noise equal to zero. However, among these 27 documents there are documents in which the percentage of useful information on a given inquiry is small with respect to the complete information of the document. An example is a report on 144 pages with retrieval number 06215, where the question is experiments on a reactor with liquid-metal fuel, and in the descriptor description there is the descriptor "zirconium."

Two documents are lost. The cause of the loss of these documents is analogous to that which was described in examining inquiry 12011, i.e., different forms of recording in the document and in the inquiry of two identical relations of ideas. Thus, in document No. 09092 in the descriptor description there is shown the descriptor "thermo-diffusion" and in document No. 09268 there is shown the descriptor "liberation"; in the inquiry there appears only the descriptor "diffusion." Thus, during modification of the inquiry inaccuracy is allowed, and the descriptors shown in the documents "thermodiffusion" and "liberation" must appear in modifications of the inquiry.

For analogous reason on inquiry 25005 there is lost the document with exploration number 06540: "processes of separation" - "waste treatment," "radioactive waste" - "liquid waste" - various forms of recording of two ideas close in meaning.

Conclusions

There is developed a system of automatic discribution of information by constant (on duty) inquiries of users (developers and scientists) from a constantly supplemented fund of information documents (reports of foreign centers and firms and dometic scientific research organizations).

The basis of treatment of information documents is a descriptor dictionary on subjects of the branch, containing 4000 basic terms and ideas.

The retrieval patterns of the parts of the document isolated in meaning (information blocks) are transferred on information maps, bearing besides a list of descriptors universal addresses and retrieval numbers and a bibliographical description and annotation (abstract) of the corresponding part of the document.

To decrease of noise during retrieval into the descriptor description there are introduced grammar elements (separating of object and theme and definition). To decrease losses special attention is payed to the composition of inquiries. The inquiry formed by the user is divided into subinquiries taking into consideration necessary hierarchical and associative links.

The descriptors of the dictionary are united in semantic groups — semantic fields allowing replacement of synonyms, terminological groups, and the idea above, below and next to. For further more precise definition of inquiries there is used feedback with the user (coupons and personal conversations).

Retrieval is realized with the help of the "Minsk-22" on the basis of rent of machine time by way of direct comparison of the descriptors of the set of subinquiries introduced into memory with descriptors of a portion of documents introduced into the machine.

Results of field testing of the system showed that noise accounts for about 70% during insignificant losses.

During further finishing of the system provision is made for strengthening grammar and introducing weight categories of individual descriptors into inquiries.

Bibliography

1. Sistema differentsirovannogo raspredeleniya informatsii (System of differentiational distribution of information). "Amer. Docum", July 1965, 16, 3.

2. Strategiya effektivnogo pochska informatsii pri "plokhom" indeksirovanii (Strategy of effective retrieval information during "bad" indexing). "Amer. Docum", July 1964.

3. Kalinin V. F. Ob uluchshenii organizatsii nauchno-tekhnicheskoy informatsii i propagandy v otrasli. (Sbornik materialov otraslevogo soveshchaniya rabotnikov organov nauchno-tekhnicheskoy informatsii i propagandy 1-2 marta) Improvement of the organization of scientific and technical information and propaganda in the branch. (Collection of materials of a branch conference of workers of organs of scientific and technical information and propaganda 1-2 March). Vyp. 2, M., 1966, 3-29.

4. Chernyabskiy V. S., Lakhuti D. G., Bernshteyn E. S. Avtomaticheskiye deskriptornyye poiskovyye sistemy s fiksirovannymi svyazyami deskriptorov (Automatic descriptor retrieval systems with fixed couplings of descriptors). In the book SEV (Trudy simpoziuma). M., 1966, 329.

5. Vanke I. Sistema mekhanizirovannogo ponska i rasprostraneniya informatsii dlya informatsioñnykh fondov i otdelov sredney velichiny (A system of mechanized retrieval and propagation of information for information funds and divisions of average value). In the book SEV (Trudy simpoziuma). M., 1966, 58.

6. Sistema khraneniya i ponska periodicheskoy literatury (CODEN). Spetsial'noye tekhnicheskoye izdaniye No. 329 amerikanskogo obshchestva po ispytaniyu materialov (v pechati) (A system of storage and retrieval of periodical literature (CODEN). Special technical publication No. 329 of the American Society for testing of materials (in print).

7. Sistema obobshchennoy obrabotki informatsii (GIPSY). (v pechati) A system of generalized data processing (GIPSY). (in print).

8. Meyer-Ulenrid. Problema avtomatizirovannoy obrabotki dokumental'noy informatsii (The problem of automatic processing of documentary information). "NTI", 1963, No. 11, 15-21.

APPLICATION OF DOMESTIC PUNCH-CARD MACHINES AND THE "MINSK-22" FOR COMPOSITION OF A DESCRIPTOR INDEX

A. I. Nadtochiy, L. N. Krasovskaya, L. P. Martynov,
and N. N. Mikheyev

In the work of information services at present for operational information and information retrieval they find wide application indices of information materials composed on the basis of key words of context of the title ("key words in context" system KWIC), and indices on the basis of key words outside the context (system KWOC). The latter form of indices started to be applied abroad relatively recently, for the purpose of increasing the effectiveness of application of indices [1]. Indices of the KWOC type are considerably more convenient in work than indices of the KWIC type, since they facilitate examination and retrieval of sources of information included in the index.

Indices of a similar type permit considerably reducing periods of bringing information to users makes it possible to detect documents from input data — key words or descriptors — more rapidly.

Indices of the type of key words found wide application in foreign atomic-energy centers. For example, Physindex [7] is the index of information materials put out by the Commissariat on atomic energy of France, system Gipsy [8], developed by the International agency on atomic energy, and others.

Such indices are used for various information materials (list of dissertations in National laboratory in Oakridge (the United States) and others) [9].

During solution of the problem of mechanized data processing and index of the descriptor type was selected.

Selection of a descriptor-type index similar in form to KWOC was caused by the available program of creation of a fund in inverted recording for IPS, realized at the TsBNTI on the "Minsk-22." Furthermore, a descriptor-type index as compared to an index composed on the basis of key words of the title of documents is based on deep indexing of documents by descriptors and, as one should expect, should ensure high probability of finding necessary information materials.

Besides what has been mentioned, during selection of an index of the descriptor type there was taken into consideration the fact that indices reduce consumption of paper as compared to card files and improve access to the fund during retrieval.

The General Flow Chart of Composition of a Descriptor Index

A descriptor index is composed in the following sequence. On every document there is composed an information card of the abstract type. To be more exact there is used an information card composed for the "SARI-1" information-retrieval system exploited in the Central Bureau of Scientific and Technical Information on Atomic Science and Technology.

A report on this system was presented at the present conference. The information card contains a retrieval pattern in the form of a number of descriptors and the bibliographical characteristic of the document (name, author, volume, number, page, location of document, etc.). Primary documents are processed using a dictionary of descriptors developed at the TsBNTI. Information cards composed from available materials go to punch. Punching — transfer of data from the information card — is carried out twice.

In the first stage of works retrieval criteria are transferred to punched cards, and on the second stage (after obtaining an inverted recording on the computer) the name of the document is punched. On the first stage there are transferred to punched cards the number of the document and descriptors in the form of codes (retrieval pattern).

The whole deck of punched cards is divided into subdecks (455 punched cards each, which is explained by insufficient working-storage capacity), and they are put into the "Minsk-22."

According to the developed program in the computer there is created recording of documents from descriptors entering the document.

The inverted recording "descriptor-documents" is created according to the following algorithm:

1. The putting of punched cards into the computer and the recording of information about documents in the form: after every document number descriptors are located sequentially. Recording is in the binary decimal system. For every document or descriptor number there is an individual working-storage cell.

2. The regulating of descriptors inside documents and the creation of a table of initial addresses of location of documents in working storage.

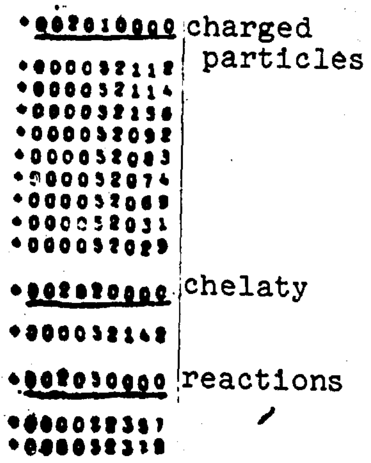
3. Creation of a single list of descriptors on all documents.

4. Formation of inverted recording of documents (on every descriptor there are written out all numbers of documents in which it is encountered).

5. Printing of obtained recording "descriptor-documents" on paper tape.

The work of the program in terms of creation of an inverted recording in many respects is similar to the work of the program of inverted retrieval on the "Ural-2" described earlier [2].

The form of the recording "descriptor-documents" on paper tape obtained from the "Minsk-22" is shown in Fig. 1, where the number +002010000 designates the code of the descriptor "charged particles (0201), and +000052112 designates the code of the reference number of the document (52112). The descriptor "chelaty" [no translation found] is code 0202, and the descriptor "reaction" is code 0205.



```
+002010000 charged
            particles
+000052112
+000052114
+000052156
+000052092
+000052083
+000052074
+000052069
+000052031
+000052029

+002020000 chelaty

+000052148

+002030000 reactions

+000052387
+000052318
```

Fig. 1. Form of inverted recording "descriptor-documents" on paper tape.

According to the obtained recording from the fund there are splintered information cards by which there is punched the name of the documents according to the developed mock-up shown in Fig. 2.

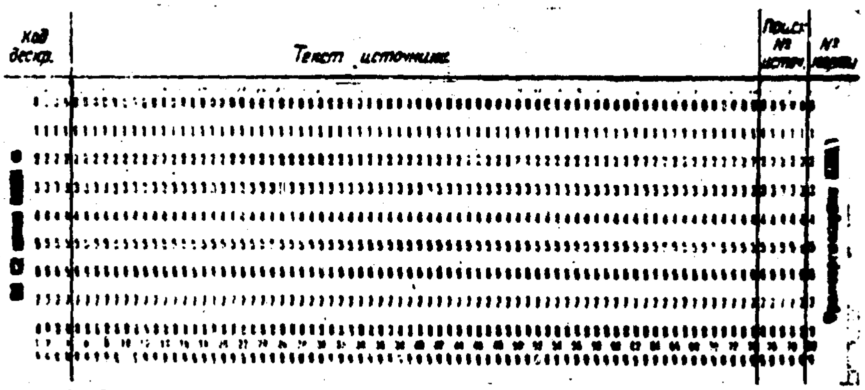


Fig. 2. Mock-up of punched card for composition of descriptor indicator.

As can be seen from the figure the punched card carries:

1. The descriptor code — four-digit number (columns 1-4).

2. The text of the document name (columns 5-74).

3. The retrieval number of the document (columns 75-79).

4. The number of the punched card (column 80).

During the composition of the index eight-digit code is broken up into two four-digit codes, and the basis is the first descriptor.

The main part of the punched card is for texts of documents. Since the text of the title will not fit on one punched card, provision is made for cards with continuation, i.e., the text of the name of documents is printed on several punched cards, which are interconnected by the number of the punched card.

During the punching of names of documents great difficulties are encountered in designations. An alphameric tabulator does not have many signs, sugh as /, %, and others, and furthermore, many formulas are not standardized in names, and therefore, in those cases in which verbal recording was unsuccessful, the formulas were manually inscribed in a tabulogram.

The following stage of work is selection of punched cards from the inverted list in alphabetic order. Documents, belonging to descriptors are taken. The sampling is done on an C-80-5M sorter. Sorter is by numbers of documents and punched cards (columns 75-80).

At present the name of the document is punched once. Therefore, if the document is encountered in the inverted list several times — and this is normal since the number of descriptors describing the document averages 8-12, then it is necessary to do extensive work on sorting the punched cards.

In further work it is proposed to improve available technology by duplicating punched cards of identical documents, punching in columns 1-4 the numbers of descriptors belonging to these documents for the purpose of sampling the necessary cards.

For text printing in the indicator of descriptors according to their codes there is developed a mock-up of the punched card different from the mock-up of printing the name of the document.

The whole field of the punched card, except for the first four columns, is for the text of the descriptor. The name of the descriptor is printed according to the inverted list mentioned at the beginning of the article.

The first four columns are omitted for displacement of the recording of the descriptor on the tabulogram - for the best reading of the indicator.

After selection of punched cards of names of documents and punched cards on which name of descriptors are recorded, on an alphameric tabulator there is printed a mock-up - a form from which the descriptor index is printed with a 1:0.7 decrease.

Figure 3 shows the page of a descriptor index. In the left part of the page there are descriptors, and in the right part numbers (retrieval) of information cards.

The index is reproduced by means of an operational polygraph. The form for reproduction is prepared on an "ERA."

In the first, experimental indicator there are included 1000 documents composed on the basis of abstracts of reports on experimental works, carried out by organizations of the State Committee on use of atomic energy.

The purpose of the index is to reduce the time of getting indicative information to the user and make it easier to retrieve necessary literature.

From a descriptor or combination of descriptors the user finds the necessary document number. After appraisal it can through the

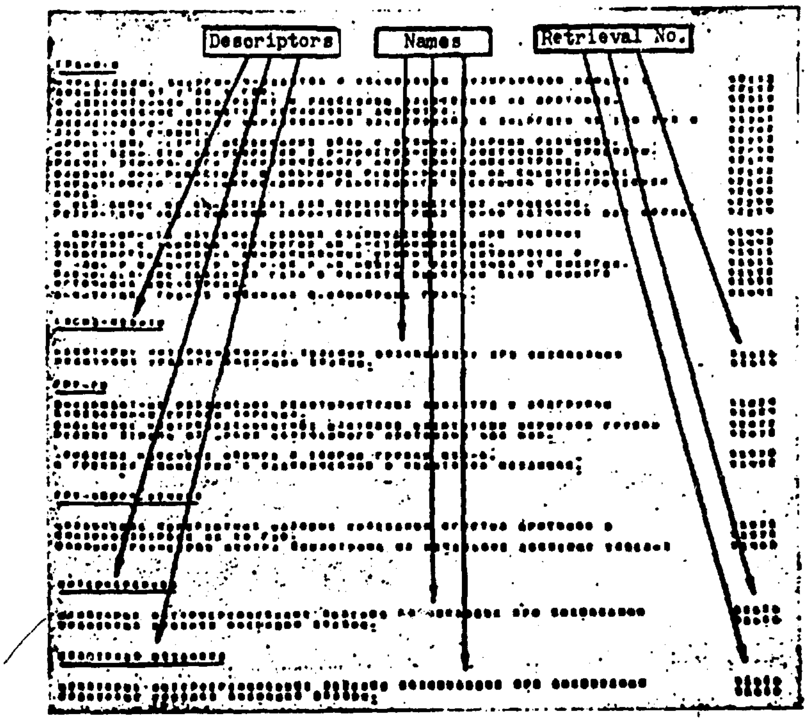


Fig. 3. Form of page of descriptor index.

reference and information fund of the TsBNTI obtain an information card, original, or copy. Certain sources of information are issued in the form of microcopies.

Conclusion

To speed up getting documents to users at the TsBNTI there is developed a descriptor index. The index is intended both for users working directly in the field of atomic science and technology and for users of other departments for which information of a similar form is needed.

Output of descriptor indices is one of the steps in the creation of an effective system of information service.

Existing computerized IPS at present can service only a certain circle of users governed by the possibilities of service of information and economic considerations.

The descriptor index should allow more widely using the existing data-processing system and giving to a wide circle of consumers retrieval by index and use of information materials available in the reference and information fund of the TsBNTI on the use of atomic energy.

Practice of work shows that the time of output of the index with the help of computers and punch-card machines can be considerably reduced as compared to such time for an index prepared completely manually.

Wide application of indices of the desc iptor type to various information materials - reports, journals, and technical specifications and records, will allow improving information service of information users.

Bibliography

1. U. Fischer. The Kwic Iudex Concept: A Retrospective View. "American Documentation", 1966, 17, No. 2, 57-70.

2. Greysukh V. L., Mikheyeva N. I., Nadtochiy A. I. Opyt realizatsii IPS inertirovannogo tipa na ETsVM (Experiment of realization of computer IPS of the inverted type). "NTI", 1965, No. 3, 21-26.

3. Lyupnits F. Ye. Sostavleniye ukazateley s pomoshch'yu alfavitnykh schetno-perforatsionnykh mashin (Composition of indices with the help of alphabetic punch-card machines). In the collection "Kompleksnaya mekhanizatsiya i avtomatizatsiya protsessov obrabotki, vydachi i peredachi na rasstoyaniye nauchno-tekhnicheskoy informatsii" (Trudy simpoziuma). M., 1965 g., 308-321.

4. Chernyy A. I., Matsak N. M., Gasanova T. G. Tekhnologiya podgotovki permutatsionnogo ukazatelya zaglaviy s pomoshch'yu alfavitnykh schetno-perforatsionnykh mashin (The technology of preparation of a permutation index of titles with the help of alphabetic punch-card computers). "NTI", 1964, No. 8, 20-26.

5. Vleduts G. E., Stokolova N. A. O sostavlenii i ispol'zovanii ukazateley tipa "klyuchevykh slov v kontekste" (The composition and use of indices of the type "key words in context"). "NTI", 1964, No. 4, 30-35.

6. Gilyarevskiy R. S., Chernyy A. I., Bibliograficheskiye ukazateli novogo tipa (Bibliographical indices of a new type). "Sovetskaya bibliografiya", 1965, No. 4, 101, 108.

7. Physindex. Sérle D. Physique nucleaire, 1966, 4-D, 7-8.

8. Gipsy Generalized Information Processing System.

9. Haas Anu Klein. Internal alerting with Keyword in context indexes. "I. Chem. Docum", 1965, 5, No. 3, 160-163; "NTI", 1966, No. 2, 1.