Information Processing & Management, Vol. 11, pp. 1–10. Pergamon Press 1975. Printed in Great Britain

METAINFORMATIONAL APPROACH TO THE THEORY OF INTEGRATED INFORMATION RETRIEVAL SYSTEMS

MAREK CIGÁNIK
Computing Research Center, UNDP, Bratislava, Czechoslovakia

(*Received* 13 *January* 1975)

Summary—Primary documents contain the retrieval information in an implicit form. The metainformational approach discussed in this paper represents some attempts to transform this implicit information structure to an explicit one. A successful solution of the problem is based on a metasyntactic analysis of texts, a creation of the semantic language in an oriented graph metastructure, and a pragmatic interpretation of metastructures based on relational contextual indexes. The metasyntactic analysis starts with a small set of inclusion metarelators, faceted relators as governing words, self-acting delimiters, and some excluding modifying phrases. A transition from the metasyntactic analysis to additional common syntactic analysis is assumed.

INTRODUCTION

MOST information retrieval systems are based on metainformation of abstracts[1]. Also, the automatic information retrieval process is fed by abstracts as surrogates[2]. The using of abstracts as surrogates is only one possibility. If we had in our mind the main criterion for evaluating the efficiency of the information retrieval system, the ability to retrieve all and only the effective items, then abstracts represent a poor resource in this respect.

The purpose of this paper is to show that there are other and richer resources of metainformation in primary documents on which we can build better information retrieval systems on a variety of integral premises. Another aim of our work is to show that it is effectively possible in newly discovered conditions to relate an external storage system to the human system in a manner that enables the combination to function most efficiently. Some other authors have shown that there are new possibilities for effective metainformation, which we can isolate by means of automatic extracting methods[3, 4, 18]. Our research is partly a continuation of these efforts. The efforts culminated in designing and achieving an operational science information system named IRIS[5].

METAINFORMATION

One branch of modern logic is concerned with the investigation of the possibilities and limits of the formal methods in logic. These enquiries into logic are also known as metalogic. The concept of metainformation is relevant to this topic. Metainformation is the grammar of the semantic information, or it is a metalanguage for information structure. In the same way as metalanguage is commonly referred to as a language used to talk about another language called the object language[6], metainformation is information about information. There are some interesting papers on the metainformational approach[7, 8]. The productive impact of metainformation as retrieval information lies on the possibility to create the theory of context based on information theory[9]. In the framework of the theory of context it is possible to create a grammar of the information structure. From this point of view all classification and/or indexing systems are metainformational systems. A very similar approach to metainformational retrieval represents the system of Farradane's relational operators[10]. This system provides nine symbols

which enable the semantic relationship between any two concepts to be explicitly specified. These concepts and symbols are: concurrence /0, self-activity /−, association /;, equivalence /=, dimensional /+, appurtenance /(, distinctness /), reaction /−, causation /:.

Naturally there is only one possible set of a metainformational relators. However, what is very important, is the hypothesis that these relations exist in the natural language texts implicitly. If it is true, then the relators language (metasublanguage) ought to exist, and this is our starting point. Now, we are going to show that it is possible to find a set of relators in the natural language form.

Let us take the relationship named by Farradane as appurtenance, and let us take the title of his paper: "Problems in Information Retrieval: Logical Jumps in the Expression of Information". It is clear by intuition that the term—Information—represents the main term in this context. The second principal term is—Information Retrieval—. Now we can say that both the terms placed in the title stand on a higher hierarchical level than other terms in the title. We have here two relators: the prepositions of, and in, respectively. We found that these prepositions and some others are hierarchical relators [11]. If we use Farradane's appurtenance relation, we can write:

Information Retrieval /(Problems in
Information /(Expression of
Information /(Expression of /(Logical Jumps.

We used here the relator /(as hierarchical relator. Its equivalents in natural language text in the above case are: of, in, :. The set of such inclusion metarelators which we have found to this time is: at, for, from, in, into, of, on, onto, to, :. The hierarchical relationship is understood here more generally than the appurtenance relationship, and from the point of view of set theory, it is better to name this relation as inclusion relation. With respect to the set theory inclusion metarelators represent the identification of the objects called ordered pairs of semantic terms. The inclusion metarelator is always between two textual strings which are carriers of semantic content. The string which has a higher degree of hierarchy is a first coordinate, and the string which has a lower degree of hierarchy is the second coordinate. If we can establish for the above prepositions the function of the inclusion metarelators, then we can divide the texts into groups and groupings respectively in the sense of Winograd's approach [12]. In his approach the WORD is the basic building block in the group, and the couplings of groups represent groupings of phrases which are used for conveying the different parts of the meaning. For example, the aforementioned title sentence has two groupings: (1) problems in information retrieval, and (2) logical jumps in the expression of information.

Expressed in another way our first partition of texts is not based on grammar, but only on a small number of metarelators as textual constants. In the same pass a dictionary of both the faceted metarelators and the modifying phrases is applied. The faceted metarelator represents a phrase which indicates some aspect (facet) of the main term. For example: the theory of, the method of, the measurement of, the analysis of, the process for, etc. [13]. The use of faceted metarelators is similar to use of word government in resolving syntactic and semantic ambiguities [14]. The modifying phrases are those types of phrases in which the inclusion relator terms have no the function of inclusion relator. For example: in our work, with respect to, in this paper, in which, in view of, point of view, for our disposal, in terms of, etc.

From prior knowledge we know that the noun groups are the information-bearers. Now we have two possibilities for determining the noun groups: (1) to find the determiners, which are the normal start for a noun group such as a, an, as, the, that, this, these, those, or (2) to find the

faceted metarelators [13], which represent an aspect to the object, which has to follow, and which diminish the vagueness of the inclusion metarelator terms. When we are using the faceted metarelators for the second partition of texts, such partition is also independent of grammar. In our example we had as faceted metarelators the phrases: problems in, expression of. Then the identification of the first object is—Information Retrieval—and the second—Information—. It is now very important to emphasize that the determiners standing in the sentence without prepositions and/or articles also have the function of metarelators.

When we have the overall semantic interpretation of discrete text, e.g. the universe of discourse, then the groupings should be the members of this universal set. We know from the set theory what such a universal set represents [15]. We shall formulate some starting axioms:

Axiom 1. We can determine the universe of discourse of a discrete document from the hierarchical context of text represented by the inclusion metarelators and/or from the organizational structure of the text. The inclusion metarelators form the contextual hierarchy of the text. The organizational structure of the text forms the hierarchy of location.

By the organizational structure of the text we understand the skeleton of document headings from the title of cover or name of the macrodocument to the skeleton of microdocument headings placed in the macrodocument. The organizational structure of text is the metainformation structure.

Axiom 2. The organizational structure of the texts is the resource of the hierarchy of location into table of contents and macroheadings. The hierarchy of location is a natural form of quasi-hierarchy of textual terms in a given universe of discourse.

Axiom 3. The groupings of text are the resource of the attribute hierarchy of semantic strings based on multiple use of the classifiers.

QUANTIFIERS OF SEMANTIC CONTENT

Axiom 4. The number of items in the organizational structure of the text is one measure of the quantity of the semantic content. The other measure is the number of ordered pairs (or groupings) in the full text isolated by means of inclusion metarelators, and faceted metarelators respectively.

QUALIFIERS OF SEMANTIC CONTENT

It is common in American documentation experiments to qualify some parts of the body of text as sources of the appropriate retrieval information, e.g. first and last paragraphs, or both. Such semantic partition of text is very rich in the case of patent literature. There are some prescribed rules of formalization in patent literature. According to such rules we can qualify the various parts of the text in the framework of the semantic categories.

Axiom 5. We can divide the knowledge contained in a discrete document into two parts: the basic knowledge and the transient knowledge. The cumulation of basic knowledge is represented by the state of the art, the cumulation of the transient knowledge is represented by the increment of knowledge, or information.

The state of the art (of knowledge or wisdom) is the cumulation of the understandable parts of knowledge or of information. The increment of knowledge is novel knowledge, or information. From the point of view of a science information system the state of the art is knowledge, but the increment of knowledge is information. In the latter terminology we need the qualifiers of knowledge and the qualifiers of information. These functions are fulfilled generally by the citation marks.

Axiom 6. Let us name an item from the organizational structure of the text with its relevant body of text as semantic complex. Then according to the allocation of the citation marks we can distinguish two types of semantic complexes: the citable semantic complex, and the noncitable semantic complex respectively. The citable semantic complex is declared if and only if the relevant body of text contains two or more citation marks of other authors than the citing authors. The semantic complex in which this condition is not true is declared as a noncitable semantic complex. The citable semantic complex is knowledge-bearing, the noncitable semantic complex is information-bearing.

According to this axiom it is possible to divide the organizational units of the text into the organizational units relevant to the state of the art in a given universe of discourse and/or relevant to information in the same universe. The verification data were gathered to this time in the fields of mathematics, information science, computer science, quantum electronics, management science, and economics.

References in a document represent the citation skeleton of special properties, but reference relations in the sentences of text represent the partition relations. It is clear that the identification of the citation marks in the body of the text is mechanical. Mr. Kessler has claimed that the bibliographic couplings represent some semantic content depending on the testing paper[16]. Now we claim that the item of the organizational structure of the text together with the sentences associated with citation marks represent the state of the art characteristics based on the citing documents. If we have a related group of items of the organizational structure with associated citing sentences, then we could speak about the semantic clusters in the state of the art in a given universe.

Axiom 7. The sentence with the reference relation represents the semantically very rich description of the problem-topic couplings. We can call this kind of knowledge the problem knowledge. The sentences associated with a reference relation sentence build the semantic foci of the state of the art in a given universe.

Axiom 8. The sentences associated with a sentence which contains a statistically significant number of both the inclusion metarelators and faceted metarelators represent the semantic foci of information.

RELATIONAL INDEXING

In correct indexing we are trying to avoid the frequent retrieval of irrelevant subject matter due to false term coordinations by partitioning the discrete topics of the universe of discourse into their various themes. All the terms relating to one particular theme are connected in some way and separated from terms relating to other themes. We can name the process of parsing the text according to the inclusion metarelators and the self-acting determiners theme partitioning. However, we must be able to distinguish interfixing from the theme partitioning. We have found that the interlocking (the connecting of terms) in the form of partitioning and interfixing is an implicit property of the organizational structure of the texts. For example let us take the latest Kochen's book from the Wiley: Becker & Hayes series. The universe of discourse is given by the name of series, e.g. *Information Sciences.* The title: "*Principles of Information Retrieval*" gives us the name of the basic theme—Information Retrieval—. This easy isolation of the basic class and basic theme of topics is based on the identification of the macrometainformation incorporated in the document. The chapters of the book represent the author's partitioning of the basic theme into themes as follows: Background to Development of Theory, The Environment of an Information Retrieval System, Information and Representation in Topics, Documentation,

Users and Uses, Topics and Directories, Specifying an Information Retrieval System, Problems and Issues.

Now we can use some rules for transforming this macroskeleton of document by:

(1) combining the faceted relators,
(2) adding the ellipses,
(3) associating the specific terms, and
(4) stopping the common words.

After algorithmic use of these rules we can present the conventional and/or relational form of representation of the document. The most appropriate form from the point of view of the relational data processing is the bracket syntax, e.g.

Information Sciences (Information Retrieval (principles of, theory development of, problems and issues)), Information Retrieval System (Environment of, specifying (Topics (Information and Representation), Documentation, Users and Uses, Topics and Directories)).

It is no problem in mapping the above structure by means of directed graphs. In such a way the relational skeleton document representation has been prepared. The relational form is very appropriate for file processing. Only the table of contents as a metainformational structure has been used, and the location metainformation has been omited.

DEPTH RELATIONAL INDEXING

The basic idea of my book on information retrieval systems [1] lies in the prior use of existing retrieval information (e.g. metainformation) and in adding further metainformation only when an acute need arises. Let us assume that some information requirements exist for "The Environment of an Information Retrieval System". This is the actual need of many information system designers. Chapter 2 of Kochen's book is relevant information answer.

According to our axioms part 2.1 of the book—Overall Perspective—has been declared as carrier of information and the remaining parts of the chapter have been declared as carriers of basic knowledge. From the indexing point of view we can prepare the relational microskeleton chapter representation as follows:

Information Retrieval System (Environment of (overall perspective, document–document coupling (citation nets), document–topic coupling (subject catalogs and indexes), document–agent coupling (authorship and readership), agent–agent coupling)).

By using axiom 8 it is possible to prepare automatically the extract of information by extracting these sentences from part 2.1:

We conceptualize *the* knowledge system *into* which *an* information retrieval system is implanted *to* consist *of* three *kinds of component parts*: (*a*) people *in* their *role as* information-processors, (*b*) documents *in* their *role as* carriers *of* information, *and* (*c*) *topics as* representations. *The number of* people, documents, *and* topics are, respectively, *the function of* time : $P(t)$, $D(t)$, and $S(t)$.

In the first sentence there are seven inclusion metarelators, five faceted metarelators, and five self-acting determiners. The second sentence has one inclusion metarelator, two faceted metarelators, and two determiners. The automatic extracting process is based on the number of weighting factors of the semantic density. The inclusion relator has weight 1, the faceted metarelator has weight 2, and the self-acting determiner has weight 1. According to this rule the first sentence has the weighting factor 22, and the second sentence 7. Weighting factor 7 is in this case the lower bound for accepting the sentence as the significant one. There is another

significant sentence in the body of the extracted text, which is in a synonymous relation with the microskeleton and therefore it is rejected. Note that a new top hierarchy term has been found, e.g. the knowledge system. This term represents a very important concept in Kochen's approach, but it is placed neither into the skeleton nor into the subject index of book. It is also clear that we have at our disposal a new index chain such as:

Knowledge system (Information Retrieval System (component parts of (People as Information-processors, Documents as Information Carriers, topics as Representations))).

We have some possibilities to proceed. For example we shall use a combination of citations and key phrases. In the IRIS system the citation code has the structure as follows:

KOCHM-74POIRAN,

e.g. the first four letters of the name, one or two initials, last two digits of the publication year, and the four letters each from the first four words of the title. The last two letters are the code of document language(AN = English). The document language code in the citation code of the chapter is replaced by the chapter number. Now we can add to the first chain the additional microchains as follows:

KOCHM-74POIR2 (Documents (Document–document Coupling (Citation nets (1CARNR-34LSDSNE, 2FAIRRA65UAMIAN, 3GARFE-55CIFSAN);
(Topics (Document–topic Coupling (Subject Catalogs and Indexes (4ROGEDJ60PFC-AN, 5AUGUJG70AOSGAN, 6STEVME65SAMFAN, 7ZIMAJM-69ICK-AN, 8WELLH-72, 9KOCHM-73VOTFAN))));
(People (Document–agent Coupling (Authorship and Readership (10PRICDD61SSB-AN, 11SWANDR61IRSOAN, 12KUHNT-62SOSRAN, 13BROMDA72PIP-AN))));
(People (Agent–agent Coupling (13BROMDA72PIP-AN, (Information System, U.S. Physicists), 14ORE-0-62TOG-AN (Graph, People as Nodes), 15MILGS-67SP—AN (Intellectual Network, Political Influence), 16POOLID58SWP-AN (17WHITHC70SPFTAN, Political Influence, Contact Net))).

We shown two possibilities of the depth relational indexing of macrodocuments, see also [17].

DEPTH RELATIONAL INDEXING OF MICRODOCUMENTS

The depth relational indexing of microdocuments is practically the same as the indexing of the chapters of a book. Small differences exist. There is a gap in the hierarchy chain on the top level. For example take the paper: P. ZUNDE and V. SLAMECKA: Predictive Models of Scientific Progress. *Inform. Stor. Retr.* 1971, 7, 103. This journal has no semantic section names, so there is a gap in the hierarchy between the top term—Information Storage and Retrieval—as the name of the journal and—Scientific Progress—as a top term in the title of the paper. During the indexing practice in testing the first paragraphs of papers we found that a hierarchical annotation exists. By the hierarchical annotation we mean one sentence form the first paragraph of the body of the text, usually from the introduction, which contains a semantically higher top term as title top term. In Zunde–Slamecka's introduction there is a sentence: "The contemporary society is witnessing a phenomenal rate of development of science and technology in all areas of human activity". This sentence represents the hierarchical annotation, which after transformation gives:

Contemporary Society (Human activity in, (areas of, Science and Technology (phenomenal development rate of))).

Now we have the possibility of choice, we opted for:

SCIENCE AND TECHNOLOGY (Scientific Progress, (predictive models based on (Information Retrieval))).

Roughly speaking, the creation of hierarchical annotation for microdocuments is sometimes a bottle-neck in automatic indexing. In the society framework it should be useful to allocate this problem to the editorial processes. However, this problem is solved in many journals. Naturally there are other approaches to the solution of this problem.

Let us proceed in our creation of the relational chains based on the microskeleton of the paper. The situation is very simple because we have:

Science Development (Markov Chain Model, Citation Data (Parameters derived from), Social Science (Social Science Disciplines, Scientific Inquiry in)).

Specifically, let us now consider the partition of knowledge. According to the above criteria the semantic complex named as Social Science (Development Trends) is a citable one, e.g. it represents the state of the art. The other semantic complexes are noncitable and therefore they represent the increment of knowledge, or information. Hence we can fulfil the relational chain of the citable semantic complex as follows:

Social Science (Development Trends in (2EARLP:69SSLUAN (Citations, Science and Technology, Social Science Disciplines (Scientific Inquiry in)))).

We can enlarge the information complexes with the relational terms too. For example:

Science Development (Markov Chain Model (Scientific Discipline Stimulators, Relative Intensity of Stimulation, First Order Markov Process), Citation Data (Parameters derived, Incitors, Complete Data Not Available)).

We can conclude that the hospitality of the relational chains is very high, and that the organizational structure of the text is extensible too.

THE INDEXING OF ABSTRACTS OR EXTRACTS

Recall and precision ratios are, according to many authors, the essential yardsticks. The very least we can do is to measure the performance of the system, for the test requests, in relation to the ideal of 100 per cent recall and 100 per cent precision. More importantly, since the members of our test user group will have indicated varying requirements for recall and precision, we can measure how far the system has been able to meet these needs. When our system is based on the metainformation structure, we have no problems in the logical adding of new items onto the system to improve recall and precision. The creation of the hierarchical annotation improves recall. An item of subject matter treated peripherally in reference relation is not appropriately covered in the indexing, but if the topic of this paper is important we can add the further material (data) to the system. At least we can add an abstract. However the exhaustivity of abstract is low. As a rule there are various types of material in the system, indexed at varying levels of exhaustivity and, therefore, we should have a data base based on the international abstracts services, in order to show the standard effect of exhaustivity levels. In such cases the automatic indexing of abstracts is very useful.

There are some techniques for automatic indexing of abstracts. We are looking for a very simple one. We can show that using the principles which have been described in this paper, it is very easy to establish the indexing algorithm. We shall try to compare our results with the indexing representations of the operational MAI system developed at the Defence Documenta-

tion Center in the U.S.A.[18]. The principal components of the MAI DDC system are two quite different dictionaries. The Recognition Dictionary consists of single words, each of which is tagged with one of 16 possible computer routines. As the text is read into the computer, one word at a time, each word is looked up in this dictionary, a designated routine is called, and a word string is built up as a possible index term. Whether or not any given string is accepted as an index term depends upon the second dictionary, the Format Dictionary. The Recognition Dictionary has over 20,000 unique words.

Our procedure is based on ten inclusion metarelators, on a small group of determiners, a small group of semantic string delimitors, and a small group of excluding phrases. This is a metaparsing procedure with the aim to identify such pairs of strings in the text which are semantic content bearers. After identifying the textual constants we are using the contextual indexing programs. These programs are creating a variety of metacatalogs or metaindexes of semantic terms[5]. For example, let us take from the appendix of Klingbiel's paper the analysed summary of Zunde–Slamecka's paper: –Progress +in science/ is essentially determined by +the stimulating effects +of information accumulation and transfer. Dynamic characteristics +of information flow/ together with +the structural properties +of flow network +in society/ should be indicative +of major thrusts +of science progress/ as well as +of its rate +of development. +A Markov chain model +of science development/ is proposed and described. It is based +on assumption that +the most recent thrusts +of scientific inquiry exert +a decisive influence +on trends of +the next immediate stage +of science development. +The application of +this model/ is demonstrated on +a sample of +the citation data +in social science. +The analysis of +this data /in terms of +the proposed model shows +a clear tendency of +the shifting emphasis +of scientific inquiry +from science and technology +to social science.–

It is clear to a programmer specialist that the resulting pattern is the result of looking-up–substitute–insert procedure. According to our approach the summary of the analyzed paper has two parts: a general statement and an author's annotation. The author's annotation is identifiable by the metaphrase "in this paper", see also another source[18]. The title of the paper is the third part of the text. In this case the general statement is the carrier of the hierarchical annotation. Now we have to find the top hierarchy term. We know that such term has to be placed in the hierarchical annotation.

Before doing so let us rap with the inclusion relators a bit more. At this time we have used these relators alone, but we must say that they stand in combinations in the sentence. Have a string: Some levels of interaction in information science. Here we have the combination of relators of and in. The top term is behind in, e.g. information science. In the analyzed paper we have such combination: the structural properties of flow network in society; the top term is—SOCIETY—. We are using the top term as an ellipse to the title according to Ranganathan's recommendation, so the new title should be: Society (Scientific Progress in (Predictive Models of)). We shall prepare now the final indexing results from the summary:

Scientific Progress (determined by (Information Accumulation (Stimulating *effects of*, *Information Flow* (Dynamic characteristics, *Flow Network* (*Structural Properties*)))));
Predictive Models (*Markov Chain Model*, *Scientific Inquiry* (Most recent *thrusts of*), Next immediate Stage, application of (*Social Science* (citation Data in (sample of)));
To (Science and Technology, Social Science (Scientific Inquiry (Shifting emphasis of (tendency of)))).

The underscored terms have been identified as the candidate index terms by MAI–DDC system, but without relations. In all cases of the Klingbiel's sample we found better results. But

this is not very important here. What is very important, is the fact that we have found the comparable data and the repeating data sets which we can process by means of the relational data base management system. Furthermore, we have very appropriate data for measuring vocabulary building in the lattice structure[19].

SUMMARY AND CONCLUSIONS

Starting from one form of relational indexing it is shown that there is a set system of relations in natural language, see also [20]. These relations are incorporated in the organizational structure of the text written in natural language. It is concluded that the text of scientific documents has an implicit faceted structure. The organizational structure of the text, the set of relations among the semantic terms as the carriers of knowledge and information, and the set of semantic strings as the coding system for information content—all represent the concept of metainformation. The effective and efficient way for building an information retrieval system is in transforming the metainformation to the retrieval information. To do so, some measuring system is needed. Some quantifiers and qualifiers of the semantic content are suggested. The proposed measure of knowledge and information is the information structure based on the state and the increment of structured semantic vocabulary. Some new principles and axioms for relational indexing have been proposed and a set of distinct semantic representations have been derived. There are many examples in the paper to demonstrate the author's approach, but we do not intend to criticize the authors cited. Some comparative discussion is planned as soon as more information on the characteristics of the IRIS system is published in English.

It was shown that this approach is similar to Winograd's and Pepinsky's one. However, the basic difference is that we are starting with the inclusion relators and faceted relators as the central metarelators and not with a predicate of the clause. The inclusion relators are surrounded on either side by semantic strings which represent the ordered pairs of the universe of discourse under study. There is another set of peripheral relators by which a deep structure of the directed graphs as measurable semantic language is created. This approach is motivated by the possibility to create as soon as possible an operating information retrieval system. This system may be a self-organizing system by introducing the additional both the tagmentic and case defined attributes of a structural language corresponding to all natural language, not only English.

We believe that the phenomena described in this paper, particularly the use of the inclusion relators, faceted relators, self-acting determiners, and qualifiers of the semantic content, will be of interest and value to those engaged in the study of the effective implementation of both automatic indexing and semantic content measuring.

REFERERENCES

[1] M. CIGANIK: *Informationsfonds in Wissenschaft, Technik und Oekonomie.* Die Wirtschaft, Berlin (1973).
[2] G. SALTON: *Automatic Information Organization and Retrieval.* McGraw-Hill, New York (1968).
[3] H. P. EDMUNDSON: New methods in automatic extracting, *J. Ass. Comput. Mach.* 1969, **16**, 264–286.
[4] C. A. MONTGOMERY: Linguistics and information science, *J. ASIS* 1972, **23**, 195–218.
[5] Computing Research Center, UNDP: Integrated Information Retrieval System. Reference Manual (in Slovak). CRC, Bratislava (1974).
[6] Z. HARRIS: *Mathematical Structure of Language.* Interscience Publishers, New York (1968).
[7] W. GOFFMAN, J. VERHOEFF and J. BELZER: Use of meta-language in information retrieval systems. *Am. Docum.* 1964, **15**, 14–22.
[8] Yu.A. CHREYDER: Information and metainformation in the thesaurus. *Nauchno-tekh. Inf.* 1974, Ser. 2, No. 4 (in Russian).
[9] M. CIGANIK: *Theoretical and Performance Properties of Context,* Martin, Matica slovenská (in Slovak) (1974).
[10] J. FARRADANE, J. M. RUSSELL and P. A. YATES-MERCER: Problems in information retrieval: logical jumps in the expression of information. *Inform. Stor. Retr.* 1973, **9**, 65–77.

[11] J. H. WHITE: The Methodology of sememic analysis with special application to the English preposition. *Mech. Transl.* 1964, **8**, 15–31.
[12] T. WINOGRAD: Understanding natural language, *Cognitive Psychol.* 1973, **3**, 1–191.
[13] Computing Research Center: The Textual Data Processing. Research Report, Computing Research Center, UNDP, Bratislava (1974).
[14] L. L. EARL: Use of word government in resolving syntactic and semantic ambiguities, *Inform. Stor. Retr.* 1973, **9**, 639–664.
[15] A. T. BERZTISS: *Data Structures, Theory and Practice.* Academic Press, New York (1971).
[16] J. C. DONOHUE: *Understanding Scientific Literature. A Bibliometric Approach.* MIT Press, Cambridge, Mass. (1973).
[17] M. CIGANIK: *Information Systems in Science, Technology and Economics.* Alfa, Bratislava (in Slovak) (1973).
[18] J. E. RUSH, H. P. PEPINSKY, N. M. MEARA, S. M. STRONG, J. A. VALLEY and C. E. YOUNG: A Computer Assisted Language Analysis System, Computer and Information Science Research Center, OSU-CISRC-TR-73-9, OHIO State University (1974).
[19] R. M. HAYES: The Decomposition of Vocabulary Hierarchies. In: *Mechanized Information Storage, Retrieval and Dissemination* (Edited by K. SAMELSON), pp. 160–191. Amsterdam (1968).
[20] H. B. PEPINSKY: A metalanguage for systematic research on human communication via natural language. *J. ASIS* 1974, **25**, 59–69.
[21] A-L. THARP: Using verbs to automatically determine text descriptors. *Inform. Stor. Retr.* 1973, **9**, 243–248.
[22] P. H. KLINGBIEL: A technique for machine-aided indexing. *Inform. Stor. Retr.* 1973, **9**, 477–494.