
**Lucyna Anna Bielicka; Janusz Paciejewski;
Eugeniusz Scibor
Institute of Scientific,
Technical and Economic Information, Warsaw**

Classification and Indexing Languages in Poland (1974–1986). Pt. I

Bielicka, L.A., Paciejewski, J., Scibor, E.: **Classification and indexing languages in Poland (1974–1986). Pt. I.**

Int. Classif. 14 (1987) No. 1, p. 23–28

The report describes the state of research and development in seven chapters, the first three of which treat the following topics: 1. General problems of indexing languages (ILs): Metainformation, metalanguage, elements of linguistics, elements of the theory of ILs, typology of ILs, tasks and construction of ILs, compatibility of different types of ILs and their linking, ILs used in Poland. 2. Structure of indexing languages: Content plane and expression plane, synonymy in ILs, pre-coordination and post-coordination, syntagmatic relationships, typology of relationships, transformation rules. 3. Indexing and information retrieval: Methods of indexing, selection of keywords in the text, relative indexing, linguistic indexing, search request formulation, weights. (Part II will be published in Int. Classif. 87-2).

Introduction

The present report belongs to the series of reports devoted to the state of research and development in the area of classification and indexing languages (ILs) in particular countries. It contains data about studies and research reports that were published (or elaborated in the form of typescripts) in Poland in the period 1974–1986. The information contained in the report has been collected on the ground of the following sources:

- the file of information about research projects in the field of information science which is maintained at the Institute of Scientific, Technical and Economic Information (Polish acronym IINTE) in Warsaw,
- retrospective searches which have been carried out using the System of Information on Research and Development Works (SYNABA),
- the journal "Aktualne Problemy Informacji i Dokumentacji" ("Current Problems of Information and Documentation") which is edited and published by the Centre of Scientific, Technical and Economic Information (Polish acronym CINTE) in Warsaw,
- series published by IINTE and CINTE,
- the journal "Zagadnienia Informacji Naukowej" ("Problems of Information Science"), edited and published by the Centre of Scientific Information of the Polish Academy of Sciences (Polish acronym OIN PAN),
- series published by OIN PAN,
- retrospective searches, carried out at the Central Library and Scientific Information Centre of the Polytechnic School in Wrocław.

As only Polish sources were utilized the report does not contain - except for a few cases - information about papers prepared by Polish research workers published abroad.

The aim of the authors of this report was to ensure the - as broad as possible - coverage of the Polish literature concerning ILs. Therefore not only studies and research reports in the strict sense but also some manuals and state-of-the-art reports were taken into account, especially those which present a certain material in a systematic way or which formulate some problems belonging to other (related) fields of science from the point-of-view of ILs. Some publications relating to descriptor languages and classification systems were considered because they treat works of great importance although the publications themselves do not have the rank of research reports.

In Poland, R & D activities in the field of ILs are mainly carried out by the following institutions: The Department of Theory of Indexing Languages at IINTE and OIN PAN, the General Linguistics Department of the University of Warsaw, and the Central Library and Scientific Information Centre of the Polytechnic School in Wrocław.

The report comprises information about 101 studies, research projects etc. concerning ILs that were accomplished in the period 1974–1986. The material is arranged within seven sections. The first three sections encompass information on those studies which concern all types of ILs whereas the further four sections (i.e. sections 4–7) include information about studies relating to particular types of them. Studies treating general problems (such as structure, functions, etc.) which are not confined to a definite type of IL are presented in the first three sections whereas studies dealing even with the same problems but in relation only to one definite type of IL are discussed within the sections devoted to particular types of ILs.

Detailed scope notes - concerning the contents of particular sections - are given in the abstracts preceding the published report because it was not always possible to achieve an accurate division of the material. Studies on weighted descriptors are, for instance, located in the section "Indexing" and not in the section "Descriptor languages" because the action of weighing descriptors is closely connected with the importance of the given lexical unit in the text of a document or with its importance for the user who is submitting the query, i.e. with some processes which take place when indexing and retrieving information which, however, do not deal with the structure of descriptor languages.

The compilers of this report wanted to stress the rapid growth of the number of studies and research projects in Poland. In the period 1974–1980 on the average a few studies concerning indexing languages were carried out each year - in sum, Polish authors have completed approximately forty studies and research projects during seven years. During the next period, 1981–1986, on the average more than ten studies and research projects were performed each year, i.e., approximately sixty studies during five years. It should be pointed out that studies conducted in 1986 were taken into account only to a small extent because of the lengthy publishing process. The above-mentioned growth finds expression not only in the number of studies and research projects but also in their character; theoretical studies as well as publications treating indexing languages in a complex way have become more and more numerous.

The references listed at the end of the report contain titles formulated in the original languages of the docu-

ments, i.e. almost exclusively in Polish; this should facilitate the identification of primary documents. However, for the convenience of the reader, full translations of all titles into English have been provided in the text of the report. The authors' names as well as the titles of the documents are printed in italics.

1. General Problems of Indexing Languages

In the paper "*On metainformation and metalanguage*", Bożenna Bojar (16) deals with different kinds of metalanguage and metainformation expressions that are formulated in natural language as well as in special meta-informational languages, i.e. in information retrieval languages (synonym of ILs). Particular attention should be paid to those fragments of the paper which treat metalanguage, i.e. a higher order language that serves to formulate higher order information about the language. Besides, information retrieval languages (ILs) are considered as such languages which are especially constructed in order to express metainformation within that sphere of it which concerns the informational contents of documents. In the simplest case the classes of documents are assigned to lexical units which occur in the search patterns of documents.

In the manual (textbook) "*Elements of linguistics for computer people*" (15), meant for persons dealing with ILs, the same author tries to encompass all that is common in natural languages and in ILs. The author gives up the traditional linguistic terminology on behalf of the terminology used in information science. Apart from general problems of information and metainformation the author treats the concepts of sign, feature, and function as well as the structure of language (systems: phonological, semantic, and syntactic). Statistical regularities of the text are also discussed. Particular attention is paid to the problem of representation (projection) and to paradigmatic relationships (i.e. to relationships between the denotations of terms) which are important from the point-of-view of the construction of thesauri.

The book "*Elements of the theory of indexing languages*" by Olgierd Ungurian (90) is not a typical textbook for students because the author has included into it also some theoretical proposals of his own which are even - to some extent - controversial. ILs are considered as an important element of the structure of information systems and are discussed jointly with the operations accomplished with the aid of ILs in the frame-work of the systems. The following problems are presented: ILs and information retrieval systems, introduction of information into the system (indexing of information) and its searching, structure of the semantic field of ILs, systems of signs used in ILs, and typology of ILs.

The manual "*Introduction to the theory of indexing languages*" written by Lucyna Anna Bielicka and Eugeniusz Scibor (13) treats - in a comprehensive way - all the theoretical problems connected with all types of ILs. At first, ILs are characterized as a kind of artificial language and the definition of ILs is formulated on the basis of the analysis of their functions and features. The levels of units used in ILs are presented as well as the essential types of IL-grammars. The semantic field of ILs is discussed also and particular attention is paid to vocabulary relationships (i.e. relationships occurring within the vocabulary of

an indexing language). Further chapters concern the stages of the introduction of information into the system (i.e. the characterization of the contents of a document and the translation of the contents into an IL). In the last chapters the application of recall and relevance factors in evaluating ILs is treated as well as the review of selected investigations relating to the efficiency of ILs.

The same authors have also elaborated another manual entitled "*Indexing languages. Their kinds and applications in information activities*" (12). In this manual, basic types of ILs are discussed. The first chapter presents a typology of ILs and reviews briefly the main types of these languages, namely: hierarchic (enumerative) classification systems, faceted classification systems, subject headings, keywords, semantic codes and other ILs having widely developed devices in order to express syntagmatic relationships. The remaining chapters are devoted to a more detailed discussion of selected ILs; these are: broad classification systems, UDC and descriptor languages. As far as broad classification systems are concerned, particular attention is paid to the Polish Subject-field Classification, elaborated in order to fulfil the role of a roof indexing language for the Polish National System of Scientific, Technical and Organizational Information (Polish acronym SINTO). In the third chapter, history, structure, and use of the UDC are discussed; special attention is paid to methods using UDC - this language being the one most widely used in Poland, being a hierarchic classification system with isolated auxiliary divisions. In the fourth chapter, the structure of a thesaurus and the grammar of a descriptor language as well as the principles of thesauri building are presented. A short information about the state of activities in the field of descriptor languages in Poland is given.

Assuming that the elaboration of a rational typology of ILs is not only of theoretical importance but also of practical value, Eugeniusz Scibor prepared a dissertation on "*Structural typology of indexing languages*" (84). This dissertation - qualifying for assistant-professorship - was defended in 1983 at the Faculty of Neophilology of the University of Warsaw. The dissertation consists of three main parts. In the first part the indexing language is characterized as a kind of artificial language; basic terms concerning ILs are explained. The second part contains the critical analysis of previously elaborated typologies of ILs. Subsequently, the following proposals of typologies are discussed:

- "descriptive continuum" of F.Jonker,
- typology of ILs proposed by J.C.Gardin,
- semantic classification of ILs proposed by E.Skorochod'ko,
- typology of ILs formulated by A.I.Chernyj,
- classification of ILs proposed by D.J.Campbell,
- typology of ILs suggested by Y.Courrier,
- typology of ILs proposed by G.van Slype,
- typologic proposals of V.A.Moskovich
- "typologic classification" of the semantic structure of thesauri, elaborated by V.D.Sidorchenko,
- set of distinctive features of ILs, collected by K.Hoppe and F.Levy.

In the third part of Scibor's dissertation a structural typology is proposed, built up by combining features connected with particular elements of the structure of an indexing language, i.e. with the vocabulary, with the presentation of the vocabulary, with the grammar and with the conversion of a natural language into an indexing language. By reducing the number of types corresponding

to the given elements of the structure of an indexing language a final typology was obtained consisting of 174 types of ILs.

The same topic has been presented by E. Scibor in a condensed form in the article *"An attempt to build up a typology of indexing languages"* (82).

Barbara Sosinska - in the paper *"The representation of knowledge in documentary information systems"* (73) - considers the indexing language as a tool for shaping the representation of knowledge in documentary information systems. She pays attention to the fact that the representation of the structure of knowledge may be incompatible with the character of information needs, which can be one of the causes of a low value of searching pertinence. The author distinguishes arbitrary and individualized representations of knowledge in information retrieval systems with access in natural language and in documentary information systems. In addition, the author discusses the functions of traditional ILs as tools for shaping arbitrary representations of knowledge and the functions of semantic networks of texts as instruments for forming individualized representations of knowledge.

In the article *"External conditions of constructing information languages"*, Elzbieta Artowicz (4) pays attention to some methodological problems facing the designers of ILs for documentary and bibliographic information systems. The dependence between the functions of the systems and their structure is shown on the one hand and the choice of concrete solutions in the conception and the structure of the language on the other hand. The analysis of the external conditions of the process of constructing an indexing language is considered as a preamble to the discussion of the stages of designing such a language. These stages are characterized by the above-cited author in the article *"Stages of creating the information retrieval language for a computerized information system"* (1). In this article an attempt is made to present a functional characterization of the process of creating an indexing language; two stages are distinguished: the stage of designing the language understood as the formulation of the conception of this language (formulation of a model) and the stage of constructing the language as the concrete implementation of the model.

In the paper *"Organization of information set as the element of the language of an information-retrieval system"*, Ewa Chmielewska-Gorczyca (18) criticizes the limitation of the grammar of an indexing language only to the rules concerning the combination of words in order to form higher order lexical units, i.e., sentences of an indexing language. Grammar, when interpreted in such a way, encompasses only problems connected with the structure of single records and does not take into account the structure of the set within which the records constitute a coherent construction. In the author's opinion the organization of the set and of the file should be considered as an element of the grammar of ILs.

In the paper *"Method of optimization of the chain structure of a vocabulary"*, Czeslaw Piekut (55) presents a method of constructing the catchword code of a term which ensures a regular distribution of the value of a term within a given vocabulary. This enables the optimum repartition of this vocabulary in the form of a chain structure. The above method utilizes the distribution of the values of signs in those terms which belong to the given vocabulary.

The rapid development of information exchange as well as of closer connections between Polish and foreign information systems has raised an immediate interest in the problem of compatibility of different ILs. Coping with this problem, Lucyna Anna Bielicka, Janusz Paciejewski and Eugeniusz Scibor have accomplished a study on the *"Concept of compatibility and methods of reaching it in indexing languages"* (11). The study is the result of the analysis of literature relating to the subject; it is also based on practical solutions utilized. A many-sided analysis of the concept "compatibility" is carried on with particular attention to its use in the field of information systems and especially in the area of ILs. Methods and techniques of reaching compatibility between ILs are described with different or similar structure, i.e. the compatibility matrix proposed by I. Dahlberg. Two essential methods of reaching compatibility between ILs are distinguished: 1) the method of direct (organic) compatibility and 2) utilization of special intermediate languages, most often limited to concordances between the vocabularies of two ILs.

In the earlier study on the *"Methods of linking different indexing languages (with particular attention to UDC and descriptor languages)"*, Eugeniusz Scibor (78) a.o. analyses selected ILs being a combination of a classification system and of a descriptor language. The author more widely discusses the problem of transformation of UDC in an indexing language with a vocabulary which meets the requirements for thesauri. Selected cases of the combined use of two ILs (e.g. of UDC and of a language based on keywords) are presented. Some conclusions concerning the usefulness and the extent of linking ILs used by Polish information establishments are formulated.

The language barrier caused by the use of different ILs can be considered as the main hindrance impeding the utilization of the collections existing in the framework of information systems. The article of Elzbieta Artowicz *"Possibilities of integrating information retrieval languages exemplified by descriptor and subject-heading languages"* (3) is devoted to the criteria of translatability of ILs and to the possibilities of integrating the descriptor and the subject-heading languages in the field of science. Also presented are the postulates concerning further development of a subject-heading language in connection with a descriptor language.

The article of Wieslaw Szczepanek is devoted to similar problems treating the *"Choice of method of linking information retrieval languages using as an example a keyword language and a thesaurus"* (76).

The study *"Methodology and organization of establishing a system of thesauri for the network of scientific, technical and economic information"* - written by Mikolaj Poletylo and Lucyna Anna Bielicka (57) - presents conception and way of establishing a thesaurus system for the Polish national information system. In the authors' opinion the national thesaurus system should consist of: 1) specialized thesauri for narrow subject-fields; 2) thesauri for branch information subsystems; and 3) a macrothesaurus (general thesaurus).

The above conception of a thesaurus system is criticized in Olgierd Ungurian's article *"Possibilities of the choice of an indexing language for the national information system"* (93). The author considers the feasibility and the purposefulness of the national system of thesauri.

He presents a number of difficulties connected with the realization of such a system, a.o. its high costs.

The survey *"Analysis of indexing languages used in the network of scientific, technical and scientific information"* was carried out by Eleonora Jabrzemska (42) in 1984 in order to collect information about the use of various kinds of ILs in Polish establishments of scientific, technical and economic information. Questionnaires had been sent to 237 establishments; 166 replies (70%) were received. The survey shows that various ILs are used in Poland; apart from languages appearing in the two-level system of ILs which is planned for the national System of Scientific, Technical and Organizational Information (i.e. apart from the Polish Subject-field Classification, descriptor languages, keyword languages and special classification systems) other ILs are widely used, especially the UDC and the subject-heading languages.

2. Structure of Indexing Languages

In the article entitled *"Structure of an information retrieval language"*, Ewa Chmielewska-Gorczyca (25) presents four basic levels of an indexing language: the level of letters (alphabet), the level of words (vocabulary), the level of sentences, and the level of texts. The author treats also the grammar of an indexing language. She distinguishes autosyntactic units (which make sentences and texts independently) and synsyntactic ones (which occur only in conjunction with other units). Also characterized are types and forms of lexical exponents of the constitutiveness of a sentence, types and forms of lexical exponents of the attachment to a sentence, and connectors (connecting signs) used in an indexing language (sentence and intersentence connectors).

The doctoral thesis of Barbara Sosinska is devoted to *"Relationships between the content plane and the expression plane in information retrieval languages"* (72). In this thesis, for the first time and in a comprehensive way, the problem of relationships between the content plane and the expression plane in ILs meant for documentary information systems is presented. In discussing the structure of language planes the author took as her starting-point the theory of Louis Hjelmslev. On the basis of the works of K.Adjukiewicz and R.Carnap an attempt was made to describe in a formal way the meaning of expressions of an indexing language using categories of denotation (extension) and connotation. The author distinguishes between direct denotation and connotation of expressions of an indexing language on the one hand and the indirect denotation and connotation of these expressions on the other hand. The organization of the content plane in ILs is presented from the point-of-view of the structure of their lexical systems. A multi-aspect typology of notational systems is built up. In the author's opinion the depreciation of the artificial notational systems on behalf of the paranatural notation (by which is meant the use in ILs of expressions taken from the natural language) is wrong because ILs with paranatural notation have a low ability to express semantic structures.

Similar problems are handled in an article of the same author which concerns the *"Structure of expressions in information retrieval languages versus the connotational and denotational interpretation of their meanings"* (74). The meaning of signs in ILs is defined in the light of the

structure of the lexical system and of grammatical rules. Two kinds of such signs are distinguished: direct designata of the signs of ILs (sets of documents) and indirect designata of these signs (physical or conceptual objects); the last ones form the subject matter of documents. ILs are divided into languages referring to the denotational or to the connotational interpretation of their expressions. Semantic codes and UDC are used as examples of the connotational interpretation of the meanings of expressions in ILs.

The article of Ewa Chmielewska-Gorczyca, *"Problems of synonymy in the theory of information retrieval languages"* (20) deals with the problem how synonymy is understood in natural languages and in ILs. A definition of synonymy in the theory of ILs - called retrieval synonymy - is suggested. Differences between vocabularies reflecting relationships of retrieval synonymy (auxiliary vocabulary) and vocabularies indicating relationships of retrieval equivalence (standardizing vocabulary) are discussed. Some methods are described for the elimination of retrieval synonymy at the level of terms of the retrieval language, at the level of phrases (synonymy of expression), and synonymy existing between a term and a phrase of the retrieval language (synonymy at different levels).

The same author in her article *"Types and terms of references in information retrieval languages"* (28) presents types and functions of references used in dictionaries of ILs (i.e. in classification schedules, thesauri, lists of subject headings, etc.). The differentiation of references into whole ("see") references and complementary ("see also") ones is illustrated by many examples. Among the complementary references she distinguishes enumerative references and textual ones.

In the next article *"Nondescriptors (forbidden terms) in retrieval languages"* (27) the above-cited author deals with one type of "see"-references, namely with forbidden terms. Various kinds of these terms are grouped within ten classes.

A subsequent article of Ewa Chmielewska-Gorczyca deals with the *"Problems of pre- and postcoordination in the theory of information retrieval languages"* (19). The author cites the views of different authors concerning pre- and postcoordination. She points out that this concept is not uniformly understood and is related to different stages of the co-ordination of the expressions of ILs. Pre-coordination corresponds to the phenomenon of parataxis in linguistics. In the author's opinion, pre-coordination is a variable feature of searching (retrieval) strategy and not a feature of an indexing language.

The doctoral thesis of Barbara Wereszczynska-Cisto treats the *"Usefulness of a detailed specification of associative relationships in the process of searching information concerning a selected branch of food technology"* (96). The thesis consists of five chapters and of an appendix. In the first chapter, associative relationships are discussed from different points-of-view. In the second chapter, two experimental thesauri are characterized (i.e. thesauri, which have been built up especially to be used in experiments): 1) a thesaurus without the specification of associative relationships; 2) a thesaurus in which such a specification is provided. In the same chapter some theoretical considerations are presented concerning character and way of differentiation of associative relationships. The third chapter contains the results of in-

vestigations which have been carried out using the two above-mentioned theoretical thesauri. As many as 122 types of associative relationships are distinguished, including 5 types of mereologic (partitive) relationships (object - constituent of the object; object - addition to the object, etc.) and 117 highly differentiated non-mereologic relationships. The fourth chapter comprises the evaluation of the usefulness of the detailed specification of associative relationships in the process of searching information concerning food technology. In the fifth (last) chapter detailed theoretical and practical conclusions are formulated. The results of the experiments show that a detailed specification of associative relationships can be considered as a tool that significantly improves recall and at the same time does not cause a decrease of relevance. The analysis of the influence of a detailed specification of associative relationships on the retrieval results show that in the situation characterized by the differentiation of queries such specification is particularly advantageous for detailed queries, i.e. queries concerning small sets of documents.

The problems presented in the second chapter of the above-mentioned thesis are also discussed in the article of the same author entitled: *"The associative relations in information retrieval languages"* (97).

The entire range of problems connected with syntagmatic relationships occurring in ILs and used in documentary information systems is presented in the doctoral thesis of Ewa Chmielewska-Gorczyca which is entitled *"Syntagmatic relationships in information retrieval languages"* (22), (23). The first (most extensive) part of the thesis is devoted to relationships appearing in search patterns of documents. The second part deals with relationships occurring in search request formulations (search specifications) whereas the third part concerns the structure of the record in a universal indexing language.

In this thesis, syntagmatic relationships are analysed which occur in the following ILs (or types of them): UDC, Colon Classification, subject-heading language, PRECIS, SYNTOL, some types of descriptor languages, Semantic Code of ASM-WRU and in the Classification of Library and Information Science elaborated by the British Classification Research Group (CRG). The analysis of syntagmatic relationships in various ILs shows the existence of similarities; therefore these relationships are mutually replaceable. The author elaborated a draft of a system which would serve the expression of syntagmatic relationships and would be common to all ILs or at least to the majority of them.

Problems similar to those presented in the above-mentioned thesis are discussed in two articles of the same author. These concern the *"Structure of an index record"* (26) and *"Linking lexical units within search patterns of documents"* (29). The first article is concerned with the syntax (physical structure) and semantics (logical structure) of the records (data formats) which are treated as sentence formulae. In the second article the types of linkages of lexical units are discussed, beginning with the simplest form, i.e. with simple co-ordination (which takes place when co-ordinate indexing is used); thereafter various kinds of links are discussed. It is shown that links can play the role of indicators of sentences of ILs or of phrases in the frame of these sentences.

The doctoral thesis of Wiesława Ogorkiewicz is devoted to the *"Influence of roles on the effectiveness of*

searching information concerning a selected branch of food technology" (52). The aim of the author was to construct a system of roles for a selected branch of food technology and to evaluate this grammatical tool as the factor for improvement of retrieval effectiveness. In the first chapter a short characterization of roles as elements of the auxiliary vocabulary of ILs is presented. The second chapter contains the description of research methods and of descriptor languages which have been built up with the aim of achieving experiments. In the same chapter theoretical considerations are presented concerning the essence of roles and the process of isolating them. The third chapter comprises the results of investigations. Particular types of roles were characterized, selected while analysing the search patterns of 500 documents; this motivated the adopted repartition of role indicators. A possible way of extending and using such role indicators is presented. Altogether 42 roles and 6 identifiers used to form complex role indicators were isolated. An analysis of the results of the investigations showed that the utilization of the elaborated set of roles in the process of retrieval has caused an average increase of the relevance factor by 16.54% and an average rise of the recall factor by 4.06 %. The fourth chapter of the thesis contains some theoretical and practical conclusions concerning the influence of the use of roles on the effectiveness of information retrieval in a selected branch of food technology. The set of roles built up by the author can be considered as a tool which significantly improves relevance and at the same time does not cause a decrease of recall.

The article *"Typology of relationships among lexical units of information retrieval languages"* - prepared by Barbara Sosinska (75) - contains a critical review of some hitherto elaborated typologies of paradigmatic relationships as well as the draft of a typology suggested by the author; this typology is an attempt to unify hitherto existing opinions. The repartition of relationships is carried out on three planes: 1) according to the types of arguments and the relationships which occur between their extensions, 2) according to the number of arguments and 3) in accordance with the possibility of using a given relationship for the organization of the vocabulary.

In the article *"Transformational rules in indexing - retrieval languages"*, Ewa Chmielewska-Gorczyca (21) - taking as a starting-point the grammatical transformation in linguistics - discusses some types of transformation that occur in ILs. Five types of transformation are distinguished: four types with a linear record (permutation, rotation, truncation, penetration) and one type characterized by the record in two lines - i.e. rolling.

3. Indexing and Information Retrieval

In the pamphlet of Lucyna Anna Bielicka, *"Conventional methods of co-ordinate indexing"* (9), existing methods of co-ordinated indexing are reviewed. From the point-of-view of criteria for keyword selection two kinds of indexing are distinguished, namely semantic indexing and faceted indexing. Semantic indexing can be used with less or more developed grammatical means (syntactic devices). Faceted indexing consists in formulating answers (on special forms) to some questions which have been prepared in advance. According to the origin of lexical

units used for the formulation of document descriptions and to the search patterns of documents it is possible to distinguish between derived and normative indexing. Derived indexing consists in formulating document descriptions and search patterns of documents with terms taken from the text and/or out of the text. Normative indexing consists in translating keywords into lexical units selected from a thesaurus or a descriptor list.

The paper "*Comparative investigation of conventional methods of co-ordinate indexing*" (6) - prepared by the same author - is devoted to comparative investigations of two kinds of indexing: free indexing and faceted indexing. The set of documents investigated contained 190 articles which were indexed four times. For the evaluation of the indexing methods use has been made of compatibility (conformability) of indexing (rule of Tanimoto). The average compatibility of faceted indexing was much higher than that of free indexing. Next to compatibility of indexing, the nature (essence) of differences between the search patterns of the same documents was studied. It was found that the big majority of lexical units which were used by only one of the indexers were linked with other units by semantic relationships. The conclusion was drawn that a low compatibility of indexing is no hindrance to obtain a high effectiveness of retrieval, particularly if the retrieval is carried out with the aid of a thesaurus or another list of lexical units provided with semantic relationships.

In the last part of the manual "*Statistical linguistics*" - written by Jadwiga Sambor (70) - applications of statistical methods to automatic indexing are briefly discussed.

In the article of Witold Marciszewski on the "*The place of keywords in the structure of the text*" (50) two kinds of text analysis are discussed which serve the definition of its structure. One of these types constitutes a parallel to the syntactic analysis carried out in phrase structure grammar while the second type comprises a certain transformational element. Both kinds of analysis are exemplified on the basis of texts relating to the science of science.

In the article "*Linguistic methods of automatic selection of indexing expressions*", Irena Szymanowska (77) describes three states of the "filtration" of text in order to select a vocabulary which is representative for a given subject-field, namely: 1) morphological analysis of text; 2) elimination from the obtained strings of these words not connected to other words within a given string; 3) selection of the most informative expressions from the strings. At the second stage two methods are used: syntactic and semantic.

In the article "*Linguistic indexing of documents by automatic methods*", Kazimierz Choros (33) discusses the analysis of natural languages performed with the aid of automatic methods.

Starting from the critical analysis of indexing methods with the aid of weighted terms, Czesław Daniłowicz, in his

report "*Problems of indexing - modelling of users profiles*" (36) presents the principles of relative indexing which consists in fixing the weights of terms on the basis of a user's profile.

Two other papers are devoted to the same problems: "*Relative indexing. Weighted descriptors and relative indexing in a document retrieval system model*", written by Kazimierz Choros and Czesław Daniłowicz (34) and "*Relative indexing on the basis of users' profile*" prepared by Czesław Daniłowicz (37). The problems of relative indexing are discussed also in two reports of Kazimierz Choros: "*Relative indexing and sufficiently ordered system answers in a fuzzy model of a document retrieval system*" (31) and "*Relative indexing in a fuzzy model of a local system of information retrieval*" (32).

In his doctoral thesis "*Problem indexing languages based on conditional interrogative sentences with one unknown quantity*", Artur Rozwadowski (64) presents the conception of a new indexing language in which natural-language interrogative sentences play the role of indexing formulae. In the thesis, the structure of the above-mentioned language as well as the conception of the construction of the standardizing dictionary are described. Complete dictionaries of prepositions, conjunctions and interrogative expressions are elaborated. The rules of problem indexing are presented and it is shown that interrogative sentences can be considered adequate tools for the purposes of indexing. It is demonstrated that the application of interrogative sentences as indexing formulae considerably enriches the means of access to a set of documents and should lead to the improvement of recall and relevance factors.

Another article of Kazimierz Choros is devoted to the problem of weighted descriptors: "*Indexing of documents with the aid of weighted descriptors*" (30). The same topic is also discussed in the article of Wojciech Seroka: "*Some problems of weighted descriptors*" (71).

In the article "*A model of document retrieval system based on the concept of semantic disjunctive normal form*", Tadeusz Radecki (61) presents a new method of retrieving documents which is based on the fundamental operators of the theory of fuzzy sets and on the concept of semantic disjunctive normal form.

In the article "*The structure of the search specification*", Ewa Chmielewska-Gorczyca (24), after having described the processes of indexing and retrieval, gives an interpretation of Boolean operators on the basis of the retrieval language. Besides this the author deals also with full-text (context) and arithmetic operators which are used in search specifications (search request formulations).

(Translated into English by E.Scibor)

References will follow at the end of Part II.