Information Processing & Management Vol. 23, No. 1, pp. 25–32, 1987
Printed in Great Britain.

0306-4573/87 $3.00 + .00

APPLICATION OF OVERLAPPING TECHNIQUE IN SELECTION OF SCIENTIFIC JOURNALS FOR A PARTICULAR DISCIPLINE—METHODOLOGICAL APPROACH

NEVENKA PRAVDIĆ* and VESNA OLUIĆ-VUKOVIĆ
Referral Centre of the University of Zagreb, P.O. Box 327, 41001 Zagreb, Croatia, Yugoslavia

(*Received* 4 *April* 1986; *in final form* 11 *July* 1986)

Abstract—A general approach to the selection of scientific journals in a particular discipline is proposed. The procedure is based on multiplicity principle; amalgamation of various methods for journal selection includes a number of data sources of different types: of international and regional origin. By application of overlapping technique combined with the ranking of the data sources according to their relative weights, the lists of nucleus journals for specific disciplines can be provided. The model operates without journal ranking.

INTRODUCTION

Academic libraries all over the world are, more seriously than ever before, confronted with the problem of profiling their periodicals collections. Their role to serve as quality filters of the world's literature with the aim to get the relevant information to the right user as fast as possible and with the minimum of cost is being endangered from two sides. First, the current exponential growth of literature is not limited to the constantly growing number of scientific journals, but refers also to the increase in their volume due to the constant rise in mean length of scientific papers. It has been recently reported[1] that the rise in mean paper length is a widespread phenomenon which started in the 1950s with the result that a typical 1980 paper is three times wordier than a typical paper from 1910–40. This affects the library storage space. On the other hand, continual increases in the subscription rates caused by rising costs of journal production affect the purchasing power of libraries. The final consequence is then cancellation of periodicals[2,3], although libraries are in fact reluctant to terminate subscriptions because they want to maintain the completeness of their journal collections.

It is not surprising that under these circumstances, in the last decade, numerous methods and techniques have been developed for collection evaluation[4] as well as for journal selection[5–8]. To profile periodicals collections as rationally and effectively as possible by selecting journals which are most suited to the users' needs is the common goal underlying most of these attempts. This is not an easy task, particularly in view of the fact that even the answer to the question "How does one define collection adequacy?" has not been reached as yet[4].

Nisonger[4] in his annotated bibliography of items relating to collection evaluation in academic libraries made a selection of the most useful contributions to this field. To mention here only some of the methods cited: coverage by abstracting services, faculty opinion and holdings of nearby libraries[9], interlibrary loan transactions[10], use studies[11,12], checking randomly selected references against the library holdings[13–15], etc. As an interesting contrast to these empirical methods, Rosenberg[16] states that an experienced and intelligent librarian is the best evaluation tool. It appears that quality-oriented approaches to the assessment of library holdings are considered superior over quantitative methods[17,18]. However, not a single one of these methods is fully adequate when used on its own, but can be helpful when supplemented by other approaches[19].

*To whom correspondence should be addressed.

And, indeed, many of the most recent studies on journal selection do combine at least two methods by applying data from different sources, e.g. citation and use data[8,20], citation analysis and readership patterns[21], two citation checking techniques[22], coverage by databases and citation analysis data[23], use and user studies[24], SDI data obtained from information service systems and the experts' opinion[7], SDI data and journal rankings according to the *Journal Citation Reports*[25].

In 1980, for the first time, a model for scientific journal selection using three different data sources (coverage by secondary services, citation data, and use data) was elaborated and offered by Dhawan *et al.*[6]. In this manner more and more sophisticated procedures are obtained. Difficulties and shortcomings of various collection evaluation techniques, which were frequently pointed out in critical analyses, are at least partially overcome by the two- or three-fold approaches. The impact of computer-aided information services on acquisition policy was considered shortly after their implementation[26], with the prediction of a direct application. The incorporation of techniques such as selective dissemination of information into the journal selection procedures has considerably improved the understanding of actual users' needs and has increased the overall reliability of the methods used. It seems appropriate to assume that the contribution of information scientists to the elaboration of these more refined methods is certainly not negligible.

In search for a better method of journal selection, our intention was to apply the experience gained to date, suggesting that complementing of data from different sources would guarantee better results. The aim was to develop a procedure that would satisfy all the requirements on a regional level. First of all, a classification of data sources according to their origin (international and local/regional) is introduced. This paper describes a methodological approach for the selection of scientific journals in a particular discipline; presentation of the general model, checking of a number of computer-aided simulations, and testing of the proposed model on the example of chemistry will be given in a separate paper[27].

MULTIPLICITY PRINCIPLE IN THE MODEL DESIGN

Concept

Three points were stated at the outset: (1) to use sources of international origin together with the data from regional sources; (2) to use as many sources of data as are available for a particular discipline and as many of them as are accessible in a given country; and (3) to avoid ranking of journals, and instead, to create one pool of journals containing input lists from all the chosen sources, and then to select nucleus journals by applying an overlapping technique combined with the ranking of the data sources.

Data Sources of International Origin: Type, Quality and Quantity

Among the available data sources, several types might be distinguished:

Type I: (i) source lists of journals in specific disciplines covered by the Institute for Scientific Information, Philadelphia, regardless of the rankings by impact factor, and (ii) data from the open literature quoting lists of core journals for specific disciplines resulting from various bibliometric studies.

Type II: (i) full lists of journals covered by selective abstracts journals, and/or (ii) full lists of journals processed by specialized databases, and/or (iii) an arbitrarily taken fraction of ranked lists of journals covered by comprehensive abstracting services (e.g. Chemical Abstracts Service).

Type III: (i) library holdings of specialized international agencies or institutions, e.g. World Health Organization, Geneva, for medicine or UNESCO Institut für Paedagogik, Hamburg, for education, and/or (ii) library holdings of leading universities or of regional/national periodicals centers.

Type IV: lists of journals for specific disciplines given in the ULRICH's International Periodicals Directory.

The quality of data provided by these sources is decreasing gradually from Type I. The source of Type IV is regarded to be of rather low value, and might be suitable only when the field of interest is not too large and when sources of higher quality are lacking.

It is suggested that the full journal coverage be taken from each of the chosen sources, with the exception of comprehensive abstracting services. The number of data sources to be included into the construction of nucleus journals lists certainly depends on their accessibility. In any case, three data sources belonging to different types seems to be the minimum. In the construction of the nucleus journals list for chemistry[27], data from five sources were used: one of Type I, two of Type II, and two sources of Type III.

Data from Regional Sources

Here again, data on relevant journals in a particular discipline could be obtained by using several techniques. The following are proposed:

(1) data on the distribution of journal articles obtained from computer-aided information services on the basis of database processing (SDI and/or retrospective searches) done for larger sets of users.
(2) data from bibliometric analyses disclosing publishing habits of the native scientists in a particular field of science.
(3) well-known methods such as: analysis of use, and/or experts' opinion, and/or users' analysis, etc.

Procedure for the Construction of the Nucleus Journals List

This procedure consists of three steps:

First, the pool of journals is formed by introducing the input list of journals from one of the chosen data sources of international or regional origin (A), followed by the input list from the second (B), the third (C), the fourth (D), and so on (up to X). The processing could be either manual or computerized. At the end of the experiment, there will be a set of journals occurring in only one of the sources, a set appearing in two of the sources (two-fold overlapping, 2×), then a group of three-fold overlapping journals, and finally a limited number of journals that are included in all the source lists (n-fold overlapping, $n\times$). Among the overlappings there are various possibilities of combinations (AB, AC, BC . . . AX, ABC . . . BCX . . . ABCX . . .). The total number of possible combinations (c) is exponentially increasing with the number of data sources which have been applied. It is expressed as

$$c = 2^n - 1$$

where n is the number of the data sources. E.g., with $n = 3$, there are seven possible combinations: 3 singles, 3 with two-fold, and 1 with three-fold overlapping journals; with $n = 7$, the total number is 127 (for details see [27]).

Upon the termination of the experiment with a given number of sources of international and regional origin, the number of journals in each of the really existing combinations is counted.

Next is the calculation of relative weights for each of the data sources used as the basis for ranking of the data sources (A–X). The criteria for ranking are based on two quantitative indicators which emerge from the distribution of the journals among the data sources. These indicators for A–X show:

I_1 – partial ratio in the total number of all existing overlappings,

I_2 – partial ratio in the reference overlappings; the upper halves of the existing overlappings are taken as reference points.

The I_2/I_1 ratio gives then corresponding relative weights. It is followed by the identification of the active combinations (a_c); active status of a combination is defined by its value, ranging from $a_{c,min}$ to $a_{c,max}$. Details are elaborated, presented and discussed in the following paper[27].

Into the group of active combinations, when determined in accordance with the above statement, regularly belong all the higher combinations (e.g. 4×–7× in a seven-sources case), then there enter some of the three- and even some of the two-fold overlappings.

As the last step in the procedure, the journals from the active combinations are picked out from the pool. The list of nucleus journals, prepared in this way, is organized alphabetically. Further ranking is not necessary, since all the journals in the nucleus are considered to be of equal value.

Should the data from an additional source be incorporated later on, due to the expected rearrangements in combinations, the calculation of the relative weights and the selection of the nucleus journals should be repeated. By the addition of data from a new source, changes always occur which eventually lead to alterations in the number of the journals in the nucleus. In shaping of the new nucleus, a certain mechanism has been recognized: (a) there is a group of journals present in the previous and in the new nucleus, (b) there is a group of journals in the new nucleus originating from the combinations that had reached the level of activity by the action of the newly added source, and (c) there is a group of journals which was present in the earlier nucleus, but due to the changes in the relative weights has now lost its activity and, consequently, is eliminated from the new nucleus. However, it could be presumed that when starting with a reasonable number of data sources, the changes are expected to be gradually diminishing.

Prerequisites for Operational Viability

The selection procedure for a periodicals collection, as described in this model is deemed to meet the needs of the scientific community in a geographically defined middle-size unit (a region, a province, a country). It is taken for granted that there exists a library-network created by cooperating libraries with a well organized interlibrary loan system. A general acceptance and understanding of the necessity for the resource sharing should certainly be well established.

DISCUSSION

The essential impetus governing the design of the model described in the preceding section was that there is a direct relation between pertinence and duplication. The experience of previous contributions to the problem of journal selection clearly shows that supplementing of data from various sources ensures a better understanding of the end-users' needs. Thus, in the three-source model by Dhawan *et al.* [6], priority is given to the journals which appear in each of the sources, to those which are cited, abstracted and used. Similarly, it was suggested earlier[28] that the list of core journals (in biomedicine) should constitute the journals with three-fold occurrence: on the list of *Index Medicus*, in *Science Citation Index*, and in the Catalog of the WHO Library, i.e. those which are abstracted, cited and included into a top-library holding. In our approach, finding an optimal number of available and accessible data sources of different types is of primary importance. Multiplicity principle which is *modus operandi* in our model might guarantee that enough reliable duplications, or better to say combinations, would occur.

The novelty introduced here is the ranking of the data sources. Having full insight into the values of the combinations, it becomes evident that, regularly, there are some of the combinations in a lower overlapping which are ranked higher than some of the combinations in the next overlapping (Fig. 1). It appears inappropriate to define the nucleus journals by the number of sources in which they are included, or by mechanical cutting between the overlappings. Determination of the active combinations provides the selection of the journals for the nucleus according to the relative weights of the data sources used in the sample. Such an approach excludes further ranking of journals as unnecessary.

Special care should be given to the choice of adequate data sources. It is obvious that the data sources are not independent. There might be, e.g., strong impact of a source of international origin either on another international source (Type I on Type III), or on a source of regional origin. This fact would not affect the outcome of the selection proce-

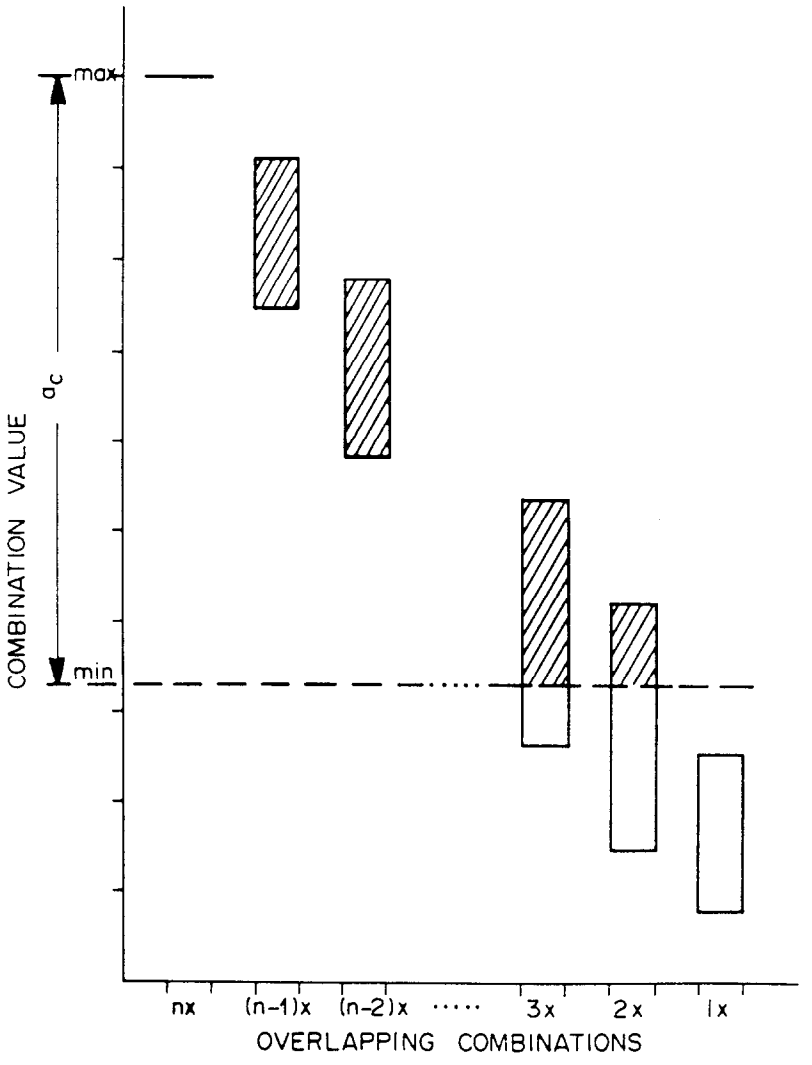

Fig. 1. Schematic presentation of the content of the active combinations (a_c) for n data sources.

dure, if the value of the combination of these two sources is below $a_{c,\min}$ (see Fig. 1). Difficulties in acquiring reliable data sources, particularly those of regional origin, are discussed later in this paper. Even if the data sources of unequal quality have to be used, this shortcoming would decrease through the processing, because the model by itself acts restrictively and selectively. All the characteristics of the model, in general, will be described together with the input/output analysis and examination of the relationships between seven data sources used in the example of chemistry[27].

The model presented in this paper seems to be rather complex. The question is whether the potential performance of the model would justify all the efforts engaged during its construction. In order to carry out at least a simplified "cost-benefit" analysis we have to turn back to the general aim of this investigation. The intention was to develop an objective and effective procedure for journal selection which would operate within the limits of the library-network in a defined region. This means that specificity of that appropriate region as a totality should be known and understood. It is obvious that, from country to country, there exist considerable differences in information-requirements mostly depending upon the needs of the on-going research and development process. Therefore, to understand the real users' needs for scientific journals on a regional scale, we introduced into the model construction a variety of data originating from regional sources. In this context, the most valuable data could be collected by computer-aided information services in the course of biblio-

graphic database processing. If either SDI or retrospective searches are done for a larger set of scientists, who are engaged on various projects in a particular discipline at different research institutions, then disseminated hits realistically reflect the selective journal needs on a regional basis. Besides, one of the advantages of this technique is that it follows and reflects the dynamics of a given field of science and identifies simultaneously the journals which have recently started publication. Namely, revealing the new journals is a special problem; in this respect, citation and use studies are completely ineffective. Last but not least, it should be emphasized that determination of scientific journals through disseminated hits is feasible during the standard database processing, and in a way it could be considered as a spin-off in the SDI process. The authors who used the SDI data in the two-fold approaches for journal selection[7,25] have also pointed out some merits of this technique.

As mentioned earlier in this paper (Data from Regional Sources), together with the use, experts, and end-users' analyses, a new technique, not yet applied in journal selection, is now proposed. It refers to the data collected from bibliometric analyses that disclose the publishing habits of the native scientists. In such analyses the relevant foreign journals, national and international, are easily recognized. The fact that scientists from one country contribute to specific journals printed abroad qualifies these journals for selection in the local library holdings. That again, if done on sufficiently large samples, might indicate selective journal needs on a regional basis.

The data we have stated here certainly belong to the categories that are not readily available in most countries, particularly not in the countries that are scientifically less developed. A major problem of these countries, as formulated by Saracevic[29], is the lack of valid and reliable facts, data, and similar hard evidence in connection with scientific and technical information, in general. On the other hand, due to their infrastructural deficiencies, these countries are seriously confronted with the problem of optimizing journal acquisition (as urgently as possible). Following these two facts, it appears that selecting nucleus journals for a specific discipline according to the procedure described in this paper, might be adequate just for smaller and scientifically less developed countries. The request for the possession of such data that will reflect journal needs on a regional basis (such as publishing habits of the active scientists, disseminated hits from computer-aided information services, citing patterns, etc.), might at the same time stimulate and direct investigations in the field of information science in that country. If this could be effectuated, the results of these investigations would serve not only in the process of scientific journal selection, but might be of much wider use (e.g. in science policy in general, research on research, scientometrics). Should this be the case, the "cost" parameter in the establishment of the nucleus journals list according to the described procedure would obviously diminish, at least partly.

Following are a few comments on the data sources of international origin. They form the backbone of the model. Four types of these data sources are differentiated and their description is given. The citation data are the first, then come the data on the coverage by abstracting services. Both of these have been frequently used in investigations with the same aim[5,6,20,30,31], and nothing new can be added to support their value and significance on the one side, or their shortcomings on the other. In Type III the holdings of regional periodicals centers are included; it is taken for comparison purposes because the predetermined frame of the model is its operability on a regional scale. The possible existence of additional sources, either international or regional, cannot be excluded. Due to the multiplicity principle, which is recommended here, data from any of the unmentioned but existing types of sources might be incorporated.

With this background, we reached one of the major characteristics of the proposed model for nucleus journals selection in a particular discipline—its flexibility. Not only that the addition of data from a new source or from a new type of sources is possible, but also the choice of the data sources to be included into the procedure is rather flexible. Regardless of the priorities, neither of the data sources is indispensable. The amalgamation process in itself, due to the overlapping technique, is expected to yield the appropriate journal selection. The choice is conditioned only by the availability of data sources for a particular discipline. The problem to be solved is how to make the relevant sources accessible. In the formulation of guidelines for the choice of the data sources, the following is summarized:

- data from international sources and data from regional sources should be combined,
- sources of international origin: the minimum is three data sources, each of different type,
- data from regional sources: the minimum is two,
- as many data sources as are available for a particular discipline and as many of them as are accessible in a given country might be used.

Another feature of the model is its feasibility. Once the data sources are chosen and input journal lists supplied, the building process may start: (1) the formation of the pool, (2) the identification of active combinations (a_c) by calculating relative weights of the data sources used, and (3) the compilation of the nucleus journals list. Although all the efforts necessary to obtain the data from the regional sources should not be underestimated, all steps incorporated into this complex procedure are feasible, provided that a well organized team satisfactorily cooperates. Academic libraries, information and referral centers should not be reluctant in tackling such a challenging problem. In fact, an evident advantage of this model is that it is founded on a "do-it-yourself" attitude. As a contrast to this concept, suggestions given recently[6] are cited: "Producing the other ranking lists at library level is time-consuming and a stupendous task. Leading indexing and abstracting organizations such as CAS, INSPEC, ISI, EI should take on to themselves the task of producing such lists and keeping them up to date. ISI and AIP have already produced such lists, and ISI updates its list, but there is urgent need for more."

We do share the opinion that producing ranking lists is a stupendous task, but, with the correction that the validity of such a statement should refer, without exceptions, to all the subjects. With this in mind we were encouraged to provide a procedure that would totally omit the ranking of journals. The result is the proposed model which is, instead, supported by the ranking of the data sources.

The applicability of the overlapping technique, based on the multiplicity principle of the methodology described here, was confirmed by a case study for the journals in the field of chemistry[27].

In conclusion, it should be emphasized that the list of nucleus journals compiled by this procedure for any of the scientific disciplines in any of the regions or countries has to be considered to contain only the necessary minimum of the relevant journals that could meet the needs of the specific scientific community at least at its minimal level.

REFERENCES

1. Trimble, V. A controllable aspect of the information explosion? Nature, 310: 542; 1984.
2. Urquhart, J.A. Why libraries are cancelling periodicals and what can be done about it. The future of publishing by scientific and technical societies. Luxembourg: Commission of the European Communities; 1978; 19-30.
3. Sanders, H.J. Troubled times for scientific journals. Chemical and Engineering News. May 30: 31-40; 1983.
4. Nisonger, T.E. An annotated bibliography of items relating to collection evaluation in academic libraries, 1969-1981. College and Research Libraries. 43(4): 300-311; 1982.
5. Singleton, A. Journal ranking and selection: A review in physics. Journal of Documentation. 32(4): 258-289; 1976.
6. Dhawan, M.; Phull, S.K.; Jain, S.P. Selection of scientific journals: A model. Journal of Documentation 36(1): 24-32; 1980.
7. Danilowicz, C.; Szarski, H. Selection of scientific journals based on the data obtained from an information service system. Information Processing and Management. 17(1): 13-19; 1981.
8. Stankus, T.; Rice, B. Handle with care: Use and citation data for science journal management. Collection Management. 4(1-2): 95-110; 1982.
9. Ash, J.; Morgan, J.E. Journal evaluation study at the University of Connecticut Health Center. Bulletin of the Medical Library Association. 65: 297-299; 1977.
10. Bolgiano, C.E.; King, M.K. Profiling a periodicals collection. College and Research Libraries. 39(2): 99-104; 1978.
11. Lancaster, F.W. Evaluation of the collection. Measurement and Evaluation of Library Services, Washington, D.C.: Information Resources Press; 1977: 165-206.
12. Wenger, C.B.; Childress, J. Journal evaluation in a large research library. Journal of the American Society for Information Science. 28(5): 293-299; 1977.
13. Lopez, M.D. A guide for beginning bibliographers. Library Resources & Technical Services. 13: 462-470; 1969.
14. McInnis, R.M. Research collections: An approach to the assessment of quality. IPLO Quarterly. 13: 13-22; 1971.

15. Nisonger, T.E. An in-depth collection evaluation at the University of Manitoba Library: A test of the Lopez method. Library Resources and Technical Services. 24(4): 329–338; 1980.
16. Rosenberg, B. Evaluation: Problems of criteria and methodology. California Librarian. 38: 17–21; 1977.
17. Strayer, M.S. A creative approach to collection evaluation. IPLO Quarterly. 13: 23–28; 1971.
18. Turner, F.L. Quality not numbers. Arkansas Libraries, 30: 8–9; 1974.
19. Evans, G.E. Collection evaluation. Developing Library Collections. Littleton, CO: Libraries Unlimited; 1979: 234–253.
20. Scales, P.A. Citation analyses as indicators of the use of serials: A comparison of ranked title lists produced by citation counting and from use data. Journal of Documentation. 32(1): 17–25; 1976.
21. Satariano, W.A. Journal use in sociology: Citation analysis versus readership patterns. Library Quarterly, 48(3): 293–300; 1978.
22. Nisonger, T.E. A test of two citation checking techniques for evaluating political science collections in university libraries. Library Resources and Technical Services, 27(2): 163–176; 1983.
23. Rose, R.F. Identifying a core collection of business periodicals for academic libraries. Collection Management. 5(1–2): 73–87; 1983.
24. Christiansen, D.E.; Davis, C.R.; Reed-Scott, J. Guide to collection evaluation through use and user studies. Library Resources and Technical Services, 27(4): 432–440; 1983.
25. Bonitz, M. Journal ranking by selective impact. New method based on SDI results and journal impact factors. Scientometrics. 7(3–6): 471–485; 1985.
26. Williams, M.E. The impact of machine-readable data bases on library and information services. Information Processing and Management, 13(2): 95–107; 1977.
27. Oluić-Vuković, V.; Pravdić, N. to be published.
28. Knežević, B. Distribution of the current medical periodical publications and its supply in the libraries of Yugoslavia (in Serbocroatian). M.A. Thesis, University of Zagreb, 1976.
29. Saracevic, T. Perception of the needs for scientific and technical information in less developed countries. Journal of Documentation. 36(3): 214–267; 1980.
30. Hirst, G. Discipline impact factors: A method for determining core journal lists. Journal of the American Society for Information Science. 29(4): 171–172; 1978.
31. Lazarev, V.S. A comparison of possibilities of various methods for the selection of scientific magazines most valuable for the specialists of a branch. Nauchno-tekhnicheskaya informatsiya, Ser. 1, (6): 27–32; 1983.