

AN ITEM-BY-ITEM SUBJECT CLASSIFICATION OF PAPERS PUBLISHED IN JOURNALS COVERED BY THE SSCI DATABASE USING REFERENCE ANALYSIS

W. GLÄNZEL,^{1,2} A. SCHUBERT,^{1,2} U. SCHOEPFLIN,^{2,3} H.-J. CZERWON²

¹*Bibliometrics Service, Library of the Hungarian Academy of Sciences,
P.O.B. 1002, H-1245 Budapest, (Hungary)*

²*Research Association for Science Communication and Information e.V. (RASCI),
Bruno-Apitz-Str. 1, D-13125 Berlin, (Germany)*

³*Max Planck Institute for the History of Science, Wilhelmstr. 44, D-10117 Berlin, (Germany)*

(Received August 31, 1999)

A serious shortcoming of bibliometric studies based on the *Social Sciences Citation Index* is the lack of a universally applicable subject classification scheme as individual papers are concerned. Moreover, the selective coverage of more than thousand scientific journals per annum proved to be an insuperable obstacle in the delimitation of social science subject areas. Subject classification of papers on the basis of assigning journals to subject categories (like those found in the various supplements of ISI databases) works well in case of fully covered and highly specialised journals in the social sciences, too, but fails for multidisciplinary and selectively covered journals. This study presents the results of an item-by-item subject classification approach, where assignment is based on the analysis of the subject categories of reference literature.

This analysis extends the results of an earlier study by the authors on the possibility of delimiting subfields in the hard and life sciences based on reference analysis. The assignment proved also reliable for a considerable share of literature in the social sciences. Due to the peculiarities of the database this share is lower in the SSCI than that in the SCI. Although an iterated application of the procedure is expected to increase the number of classifiable publications, it is suggested that in the social sciences the method should be used in combination with other means of subject assignment.

Introduction

The importance of delimiting scientific subfields for valid data processing in information science and bibliometric research has often been stressed. Several attempts have, therefore, been made to create a subject assignment for those multidisciplinary bibliographic databases which lack appropriate subject classification schemes

(e.g., *Narin*, 1976, *De Bruin and Moed*, 1993, *Braam et al.*, 1991, *Small*, 1997, 1998 and *Glänzel et al.*, 1999). Particularly, the databases *Science Citation Index* and *Social Sciences Citation Index* of the Institute for Scientific Information (ISI, Philadelphia, PA, USA) do not provide any direct subject assignment for the indexed papers. However, since the annual *Science Citations Index Guides* and the *ISI Journal Citation Reports* (JCR) contain regularly updated lists of (S)SCI journals assigned to one or more subfields (ISI Subject Categories) each, more or less modified versions of this Subject Category scheme are used in bibliometric studies to compensate the lacking subject-heading system. Thus, an indirect subject assignment of individual papers can be performed by means of the journals in which they have been published. Taking into consideration that journals are often not devoted to a single topic, the delimitation of subject areas based on journal assignment is obviously less precise than that based on subject headings of individual publications. In an earlier paper *Glänzel et al.* (1999) developed a method for delimiting the subject of papers published in multidisciplinary and general journals of the *Science Citation Index* (SCI) by applying an item-by-item subject classification scheme where assignment is based on the analysis of the subject classification of reference literature. The approach is based on the assumption that the ISI Subject Categories system is sufficiently precise for specialised journals which are devoted to one particular or two related subfields. The assignment of individual papers published in multidisciplinary and "general" journals and, consequently, needing an individual assignment was based on the analysis of their reference literature.

The subject of a large share of otherwise unassigned publications recorded in the SCI could be successfully determined with the help of the new method. The methodology proved to be efficient at both the level of major fields and subfields. It allows an iterated application which is expected to improve the efficiency of classification.

In the present study the authors attempt to apply this method to assign papers published in *selectively covered* journals of the *Social Sciences Citation Index* (SSCI).

Although the structures of the two databases are formally identical, that is, SCI and SSCI use the same bibliographic fields with coincident components, there are differences concerning the bibliographic data of journals covered by the two databases which should in any case be heeded in bibliometric studies.

The most striking peculiarity of the SSCI database refers to journal coverage. Whilst in the SCI *all* papers of given document types are recorded which have been published in a journal covered by the database, in the SSCI both *fully* and *selectively* covered journals can be found. This seemingly unimportant difference has serious consequences, first of all, on the subject classification. The assignment of papers to subject categories

defined through journals is quite impossible if the papers in question are published in selectively covered journals since, in general, only fully covered journals are assigned to subject categories. Selectively covered journals are often represented by very few publications which are in part indirectly related to the main scope of the journal (for instance, an archaeological paper concerned with ancient metallurgy, therefore published in a chemistry journal). In these cases the ISI Subject Category in the SCI – if such exists – would be misleading and should therefore not be used.

The share of papers without corporate addresses and/or without cited references is much higher in the social sciences than in the sciences (in some subfields such as library science or business this share amounts even 25–30%, cf. *Glänzel and Schoepflin, 1999*). Nevertheless the method is expected to be applicable in subject fields where the communication behaviour is similar to those in the sciences. The application of the method to the subject classification of papers concerned with subject areas such as psychology, sociology, economics and information science published in both fully and selectively covered journals is, therefore, the main target of the present study. In addition, the results in selected fields like business, political science and law are also checked. The effect that the share of papers without corporate addresses and/or without cited references is particularly large in these fields, has to be relativised by that fact, that a considerable part of such publications cannot be considered conveyors of original scientific information anyway.

A second peculiarity concerns the representation of subject areas in the SSCI. On one hand, there are joint borderlines between the social sciences and the sciences (for instance, in psychology/psychiatry), on the other hand, there is a considerable overlap between social sciences and humanities, thus, e.g., the fields archaeology and history are represented in both the SSCI and the AHCI databases. Therefore, some interesting disciplines at the borderlines between social sciences and life/hard sciences, and social sciences and humanities, respectively, have not been taken into consideration.

Methods and results

In the present investigation, seven subject fields have been selected. As already mentioned above, delimiting research fields in the social sciences is more "fuzzy" than in the hard or life sciences. Moreover, publication counts in social science journals and subfields result often in too small numbers for reliable statistical analyses of individual journals and subfields. Unlike in the sciences (see *Glänzel et al., 1999*), the authors have therefore renounced the breakdown by subfields. The subfields are clustered into larger subject categories instead and thus appropriate areas in social sciences are (re)defined.

In order to avoid biases in favour of the sciences, subfields such as clinical psychology, psychiatry and statistics have been excluded from the analysis. The description of the selected fields based on the corresponding subject categories for social science journals is given in Table 1.

Table 1
Definition of seven social-science areas based on ISI's Subject Categories

Subject area	Subfields
1. Business	Business Management
2. Economics	Business, Finance Economics
3. Law	Criminology & Penology Law
4. Political Science	Medicine, Legal International Relations Political Science Social Issues Planning & Development Urban Studies
5. Psychology	Psychology Psychology, Applied Psychology, Developmental Psychology, Educational Psychology, Experimental Psychology, Social
6. Sociology	Ethnic Studies Sociology Women's Studies Demography Family Studies Social work
7. Information & Library Science	Information Science & Library Science

Figure 1 shows the coverage structure of the 1995 volume of the SSCI and the representation of the seven fields in this volume. Here the subject classification of journals based on the areas in Table 1 has been used. Almost 50% of the papers of the document type *articles*, *letters*, *notes* and *reviews* indexed in the 1995 volume of the SSCI are covered by the selected fields. About one third of the papers could be assigned to other subject categories, and slightly more than one fifth of all papers of the above document types have been published in the by far more than thousand selectively covered journals.

Using the subject areas presented in Table 1, we can proceed from the assumption that the precision of subject classification of specialised journals, that is, of journals assigned to one or two subject categories of the above subfield list, is sufficient for bibliometric macro and meso analyses. In the second step only papers published in such journals were assigned to their subject areas. In order to validate the new method, papers published in selected specialised journals have also been reassigned.

The topic of most papers published in scientific journals is characterised by their reference list. In the third step it was attempted to identify those ISI journals in the reference lists of source publications (in the multidisciplinary and selectively covered journals) which are clearly devoted to at most two specific subjects according to the revised subfield list (cf. Table 1). Once a journal name in the reference item was successfully identified, it could then be replaced by its subfield code(s). In principle, two approaches are possible, in particular,

- (1) using thesauri, where all spelling variances of the journal abbreviations and the journal's subfield code(s) are listed,
- (2) matching cited reference and source items by an appropriate identification-key, where the source record contains besides the identification-key also information about the paper's subject assignment.

Like in the earlier study on subject classification of papers published in multidisciplinary science journals (Glänzel et al., 1999), the second method has been applied. This allows the iterated application of subject assignment in subsequent years. Thus papers published, for instance, in selectively covered journals which have once successfully and unambiguously been assigned by the new method, can be used to delimit the field of citing papers.

For better understanding we briefly recall the assignment methodology developed in the above-mentioned paper: all references made to items other than those publications in journals assigned to one or two of the seven selected subject areas have been removed. In other words, the reference lists of these papers have been reduced to those items which were published in journals which (almost) unique subject assignment according to the above criteria. The most frequent subject categories in the reference list then serve as the assigned subject categories of the paper in question. If the reduced reference list proved to be empty then the corresponding papers could not be assigned to one of the seven fields.

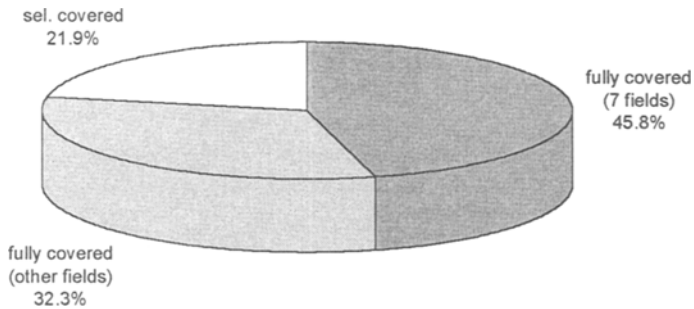


Fig. 1. The coverage profile of the 1995 volume of the SSCI

Table 2
 Example for identified papers published in *Social Science Journal* (1995)
 (F₁, F₂, F₃ subject codes with rank *i* by frequency, # absolute frequency)

F ₁	#	F ₂	#	F ₃	#	Title
4	2	0	0	0	0	The Impact of Opening Markets on Mexican Male-Female Wage and Occupational Differentials
4	2	2	1	0	0	After NAFTA – Western-Hemisphere Trade Liberalization and Alternative Paths to Integration
2	1	0	0	0	0	NAFTA and the Mexican Economic – Crisis – Causality or Coincidence
6	1	4	1	0	0	Culture as a Determinant of Reasons for Migration
4	1	0	0	0	0	Government Coercion – An Exploratory Analysis
5	1	4	1	0	0	Political Tolerance Hypotheses and White Opposition to a Martin Luther King, Holiday in Arizona
3	1	0	0	0	0	Tough Laws – Policy-Maker Perceptions and Commitment
3	2	0	0	0	0	Brains for the Bucks - School Revenue and Student Achievement in Oklahoma
6	1	0	0	0	0	The Occupational Mobility of Black Males Revisited – Does Race Matter
2	2	0	0	0	0	Threat to High Market Value Agricultural Lands from Urban Encroachment – A National and Regional Perspective

Table 2 presents an example for frequency distributions of the cited journals' subject categories in the reduced reference lists of the American multidisciplinary journal *Social Science Journal*. The most frequent three subject code have been displayed in descending order of their absolute frequency. The field codes correspond to those given in Table 1. The results substantiate the multidisciplinary character of this journal.

The new subject assignment has been applied to articles, letters, notes and reviews of the 1995 annual volume of the SSCI. All references from the period 1991-1994 have been analysed. This time period has been determined on both an empirical basis and based on the results of ageing studies of scientific literature (Glänzel and Schoepflin, 1999). On one hand, in the hard and life sciences a considerable part of references is not older than five years (Price Index), on the other hand, older references proved to be less specific in the context of subject delimitation.

Unlike results obtained from the analysis of the multidisciplinary and general journals covered by the SCI, it is difficult here to interpret the share of classifiable publications in the multidisciplinary and selectively covered journals. The main reason is that the seven selected fields are expected to represent but less than two thirds of the publications recorded in the SSCI. Moreover, the share of papers without addresses and references in the SSCI is greater than that in the SCI. This applies also to journals with otherwise unique subject assignments. Altogether 28.1% of the papers published in selectively covered journals could be assigned to one of the *selected* fields. The share of assigned papers in the groups "selected seven fields," "other fields but fully covered journals" and "selectively covered journals" are presented in Figure 2. The value in the second group is, therefore, fairly low and corresponds mainly to multidisciplinary journals.

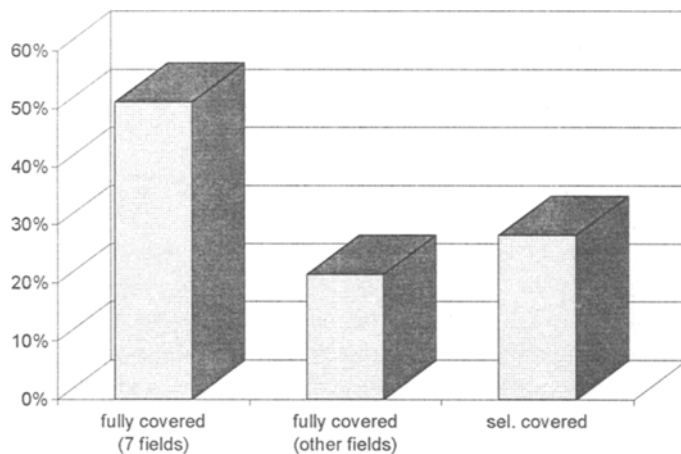


Fig. 2. The share of assigned papers by coverage profile

Figure 3 shows the distribution of publications in selectively covered journals over the seven fields. In order to obtain additive results, that is, to obtain shares which can be summed up to the total, a unique assignment has been made. In particular, papers have been assigned to the field with the greatest frequency, provided the relative frequency exceeded 50%. Otherwise the paper has been considered "interdisciplinary". Almost three quarters could not be assigned to one of the seven fields. But this large share does not necessarily imply a flaw in methodology. Manual checking showed that, for instance, papers published in chemistry journals were mainly concerned with archaeology or history of science and social sciences, that is, fields which have been excluded from the analysis.

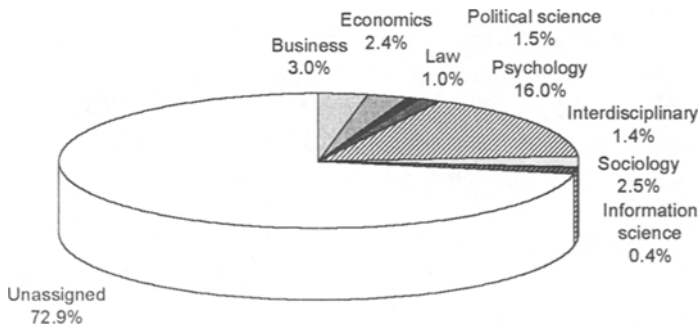


Fig. 3. The subject profile of selectively covered journals

56.9% of assigned articles were concerned with psychology. The reason for this relatively large share is first of all the citations of publication in psychology journals received from papers concerned with psychiatry and clinical psychology. If citations are regarded as information use, transfer of information in the social sciences may often have sources and targets outside the own field or even outside the (social) sciences. This particular phenomenon in the social sciences has already been pointed out by Glänzel and Schoepflin (1999). The areas sociology, business and economics are represented by about 9-10% of all assigned papers each. Since often very few papers are recorded from each selectively covered journals, an analysis by selectively covered

journals will not be presented. In the following we will analyse the publication profile of the multidisciplinary journals *Daedalus*, *Jahrbuch für Sozialwissenschaft* and *Social Science Journal* (see Figure 4).

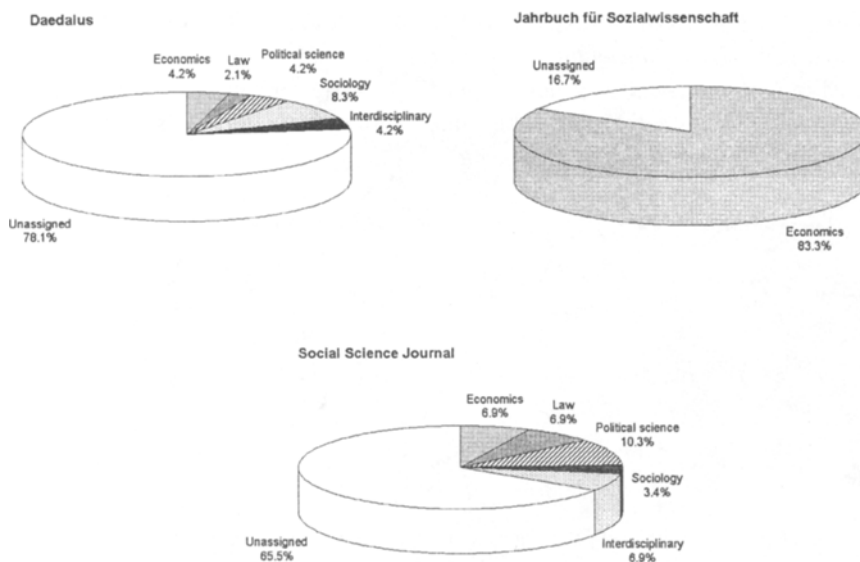


Fig. 4. Distribution of publications over subject areas (multidisciplinary journals)

Almost 80% of papers published in the journal *Daedalus* could not be assigned to any field through the reference analysis although the papers have cited a sufficiently large number of references. The subject profile of the assigned papers generally corresponds to the journal's profile. One reason for the low share of classified papers can be found in the already mentioned peculiarity of science communication in social science areas such as political science, business and law (cf. Glänzel and Schoepflin, 1999). By contrast, the "small" German journal *Jahrbuch für Sozialwissenschaft* proved to be an journal devoted to economics, 83.3% of the papers could be assigned to this area. Finally, about one third of the *Social Science Journal* could be assigned to the areas political science, business, law and in smaller part to sociology. These three examples may illustrate the possibilities, but also the limitations of the application of the item-by-item classification method to papers recorded in the SSCI.

As in case of the sciences, the method has also been applied to two specialised journals selected for validation. The method is expected to confirm the journal assignment by that of the individual papers. *American Sociological Review* (*ASR*) has been selected as a sociological journal and *Developmental Psychology* as a journal in psychology. Figure 5 shows the results. The share of unassigned papers lies between 7% (*ASR*) and 13% (*Developmental Psychology*). The overwhelming share of papers published in *Developmental Psychology* (94.1% of assigned papers) could be assigned to psychology. The other papers have either not been uniquely assigned (3.5% of classifiable papers) or have been assigned to sociology.

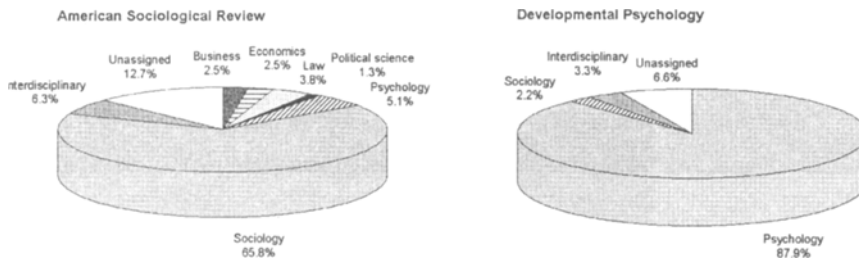


Fig. 5. Distribution of publications over subject areas ("specialised" journals)

75.4% of all assigned papers published in *ASR* have been assigned to sociology, 7.2% have not been uniquely assigned to any single field. The rest is almost uniformly distributed over the other selected fields, only information science is not represented in this journal.

In the case of specialised journals, these results are certainly influenced by journal self-citations, however, the overwhelming shares cannot be explained by journal self-citations alone. The share of journal self-citations in all cited references is in both cases below 10% (in case of *ASR* even below 5%). Thus the results corroborate the assignments of the two selected journals according to ISI's Subject Category scheme.

Conclusions

The subject of a considerable share of otherwise unassigned publications could be successfully determined. It should be stressed that the analysis was restricted to seven previously selected and redefined subject areas. All other subject categories have been

ignored. Such restriction is necessary since a considerable part of recorded publications are out of the scope of the traditional social sciences. Under this restriction, almost 30% of all papers published in 1995 in selectively covered journals could be assigned to one of the seven areas. Since there is no other possibility of computerised subject assignment for these papers, the applied method provides a first approach to a complete subject classification of SSCI publications. The method may therefore be combined with other delimitation techniques to improve its efficiency.

On the other hand, a certain share of papers recorded in the SSCI cannot be considered to report original research results. Similarly to the sciences, non-research papers (for instance, news or scientific journalism otherwise classified by ISI as research articles) should be excluded in the case of the social science studies, too.

References

- BRAAM, R.R., H.F. MOED, A.F.J. VAN RAAN, (1991) Mapping of science by combined co-citation and co-word analysis I., *JASIS*, 42 (4) 233-251.
- DE BRUIN, R.E., H.F. MOED, (1993) Delimitation of scientific subfields using cognitive words from corporate addresses in scientific publications, *Scientometrics*, 26 (1) 65-80.
- GLÄNZEL, W., U. SCHOEPFLIN, (1999) A bibliometric study of reference literature in the sciences and social sciences, *Information Processing and Management*, 35, 31-44.
- GLÄNZEL, W., A. SCHUBERT, H.-J. CZERWON, (1999) An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis, *Scientometrics*, 44, 427-439.
- NARIN, F., (1976) *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity*, Computer Horizons, Inc., Washington, D.C.
- SCHUBERT, A., W. GLÄNZEL, T. BRAUN, (1989) Scientometric Datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major fields and subfields 1981-1985. *Scientometrics*, 16 (5-6) 3-478
- SMALL, H., (1997) Update of science mapping: creating large document spaces, *Scientometrics*, 38 (2) 275-293.
- SMALL, H., (1998) A general framework for creating large-scale maps of science in two or three dimensions: The SciViz system, *Scientometrics*, 41 (1-2) 125-134.