
Book Reviews

SIMON, Hans-Reiner: **Research and Publication Trends in Systematic Zoology 1758–1970**. Berichte der Ökol. Aussenstelle Schlichtern, H. 11, Frankfurt/Main – Schlichtern: J.W. Goethe Universität, Ökol. Aussenstelle 1982. 403 p. (available from the author, Römerstr. 44, D-6084 Gernsheim)

This book is noteworthy both due to its contents and the form of representation of the material. This is a well-commented scientometrics guide-book concerning the growth of systematic zoology. The object of description is the increase in the number of species (for separate groups of animals) known at present, in the number of publications, the number of pages, and the number of original and abstracts journals. The book embraces the period between 1758 and 1970. The material is represented on 57 tables and 81 graphs. The latter are very attractive: they are large enough and are easily grasped, besides many of them include schematic images of corresponding animals, which makes their differentiation and retention of the uniform material easier. The data have been treated mathematically: all the parameters of the growth curves have been calculated. The change in the growth parameter is represented by a large number of graphs. At the end of the book an attempt is made to compare the growth of publication with that of biological knowledge proper, the latter being measured by such indices as indexing in *Zoological Records*, the number of important publications, the number of newly described species and higher taxons, the emergence of major theories. The growth of the number of publications is compared with the growth of economic resources (for the period of 1880–1913), the growth of the number of zoologists and manpower.

The problem of growth is a central one in scientometrics and perhaps in the entire science of science. The book gives a strikingly detailed and substantial study of the problem, and because of that the results of the given study acquire general scientific meaning leaving the boundaries of the branch of science they are based upon. The book will certainly attract a broad circle of researchers – historians of science, library scientists, information scientists and scientometrists proper. It is also convenient for teaching since the methodological part is written very thoroughly.

V.V. Nalimov

Prof. V.V. Nalimov
Lab. of Mathematical Theory of Experiment
Lab. Building "A", Moscow State University
Moskva 117234, USSR

GAUS, Wilhelm: **Dokumentations- und Ordnungslehre. Lehrbuch für die Theorie und Praxis des Information Retrieval**. (Documentation and Classification

Science. Textbook for Theory and Practice of Information Retrieval)

Berlin – Heidelberg – New York – Tokyo: Springer Verlag 1983. 504 p., 113 figs., DM 79,–, ISBN 3-540-12777-1

The author draws upon his copious store of practical experience and provides ample advice for the acquisition, indexing (mainly intellectual), searching for and conveying of information to the user, and for the establishment and revision of an information system. The individual chapters are illustrated by well selected examples, elucidated through enlightening questions and made accessible by an excellent index. The deficiencies of contemporary retrieval system theory are not expressly stated by the author, but they become apparent through the perspicuity of the presentation. The book is recommended to any reader who is striving for some insight into the intricacies of information retrieval and is in search of practical advice in this field.

Das treffsichere Wiederfinden von Informationen hat sich zu einem großen Arbeitsgebiet entwickelt. Viel Arbeitskraft und Geld wird weltweit in die praktische Ausübung dieser Tätigkeit investiert. Die theoretischen Grundlagen hierfür jedoch lassen zu wünschen übrig, denn andernfalls würden nicht so viele Informationssysteme aller Größenordnungen so viele Mängel an den Tag legen, und es wären nicht bereits so viele von ihnen gänzlich gescheitert. Eine gute Theorie hätte solche Fehlschläge voraussagen und verhindern können. Ein Lehrbuch, das Theorie und Praxis in enger Verbindung behandelt, weckt große Hoffnungen. Es kann sehr dazu beitragen, daß künftige Informationssysteme auf einer Grundlage eingerichtet werden, die wenigstens dem gegenwärtigen Stand der Theorie entspricht.

Eine wichtige Voraussetzung hierfür erfüllt das Buch allein schon dadurch, daß es in der Sprache des Praktikers geschrieben ist und auf der langjährigen didaktischen Erfahrung seines Verfassers aufbauen kann. In großer Vielfalt werden nützliche und praktische Ratschläge für das Beschaffen, Indexieren, Wiederaufsuchen und Weiterleiten von Informationen gegeben. Die Ausführungen darüber, wie sich die Kommunikation zwischen dem Informationsvermittler und dem Informationssuchenden gestalten sollte, verraten große praktische und psychologische Erfahrung, mag sie sich auch nicht immer unmodifiziert auf andere Verhältnisse übertragen lassen. Jedes Kapitel wird durch eine Reihe von sorgfältig zusammengestellten, lehrreichen Fragen abgeschlossen, die an anderer Stelle beantwortet werden. Vertieft und verdeutlicht wird der Lehrstoff durch viele gute Beispiele. Zwar ist das Buch eine Niederschrift von Vorträgen zur Ausbildung von medizinischen Dokumentaren, aber die Beispiele sind so gewählt, daß sie auch für einen Nichtfachmann dieses Gebietes verständlich sind.

Nützlich sind die für manche Novizen desillusionierenden Ausführungen über die notwendige Organisation der Arbeit zum Einrichten und Erproben eines Ordnungssystems, sowie auch zu dessen regelmäßiger Revision, überaus realistisch auch die Ausführungen über den hierfür notwendigen personellen Aufwand, die Anforderungen an die Qualifikationen des Personals und an die Leistungsfähigkeit der technischen Hilfsmittel. Übersichtlich und präzise wird dargestellt, welche Fragen bei der Planung eines Informationssy-

stems vorrangig zu klären sind und wo die Grenzen der technischen Hilfsmittel, wie Kartei, Liste, Register, Endziffernindex, der Handlochkarte und des Computers, liegen. Es ist ein Zeichen großer praktischer Erfahrung auch auf diesem Gebiet, wenn der Verfasser unter bestimmten Umständen, die bei persönlichen Sammlungen oftmals herrschen, von der Verwendung von EDV abrät, in anderen Fällen aber wiederum nützliche, praktische Ratschläge für den EDV-Einsatz gibt. Ein Beispiel hierfür ist das Systematisieren eines Schlagwortvokabulars, wo die Flexibilität und Effizienz des Arbeitens mit handgeschriebenen Zetteln auch durch raffinierte Techniken kaum überbietbar sein dürfte. Daß manche Vorschläge, hat man sie erst einmal empfangen, nachträglich sehr naheliegend erscheinen, mindert nicht ihren Wert, denn sie ersparen dem Neuling viel Lehrgeld. Beispielsweise sollte man nicht die Erfahrung erst selber machen müssen, daß man beim Einführen eines neuen Deskriptors in den Wortschatz das Einführungsdatum festhalten sollte, daß man bei einer jeden Dokumentationseinheit festhalten sollte, mit welchen Deskriptoren man sie indexiert hat und daß man als Informationsvermittler immer mit dem Fragesteller selbst kommunizieren sollte und nicht über nochmals zwischengeschaltete Mittelspersonen.

Die Akzente des Buches liegen zwar auf der Beschreibung und Empfehlung des gebundenen Indexierens, als dem verbreitet üblichen Verfahren. Es wird jedoch auch auf die Möglichkeiten und Grenzen der automatischen Indexierungsverfahren und der Freitextrecherche eingegangen. Die Probleme bei den beiden letztgenannten Verfahren werden allerdings in einem noch relativ günstigen Licht dargestellt. Hier haben sich wohl die Publikationen der Forschungsgruppen auf diesen Gebieten ausgewirkt, in denen fast immer die eigenen Ergebnisse sehr wohlwollend und nachsichtig im Vergleich mit den weniger stark mechanisierten Verfahren beurteilt werden. Sehr informativ sind auch die Ausführungen über die verschiedenen Varianten der Kombination von intellektuellem und maschinell Indexieren, bzw. mit der Freitextrecherche.

Was die Theorie anbetrifft, der das Buch ebenfalls gewidmet ist, so spiegelt es zwangsläufig den relativ niedrigen Entwicklungsstand derselben wider, und es wären wohl einige Worte der Skepsis in dem Buch hierüber am Platze. Der Leser und Lernende sollte nicht den Eindruck gewinnen, daß auf diesem Gebiet alles im Lot ist. Wenn beispielsweise einer verbreiteten Auffassung zufolge die Reproduzierbarkeit des Indexierens („consistency“) ein Qualitätskriterium für die Indexierung ist, warum ist dann mitunter gerade die best reproduzierbare Indexierung, nämlich die vollautomatische, die am wenigsten befriedigende? Wenn es wirklich den oft behaupteten strengen Antagonismus zwischen Relevanz und Vollständigkeit gibt, durch welche Eigenschaften erklären sich dann Informationssysteme, in denen für beides die 100%-Quote erreicht wird, wie sie beispielsweise in der Chemie geläufig sind? Auch muß man die Kurvenzüge im Relevanz-Vollständigkeitsdiagramm irgendwo vor dem Erreichen des Koordinatenkreuzes abbrechen, weil sie dort sinnlos werden. Spätestens dort

wird offenkundig, daß die ganze Regel so nicht stimmen kann. Die Aufzählung der Unklarheiten und Widersprüche, die die heute verbreiteten Theorien offen lassen, ließe sich noch fortsetzen.

In ernste Schwierigkeiten aber gerät der Verfasser dadurch, daß er von dem sehr engen und altertümlichen Klassifikationsbegriff ausgeht, wie er beispielsweise in den USA noch weitverbreitet anzutreffen ist. Nicht etwa das Zusammenbringen von verwandten Gegenständen in einem Abstraktionssystem macht hier das Wesen der Klassifikation aus, sondern in erster Linie der Umstand, daß jeweils nur ein einziger Deskriptor einer Dokumentationseinheit zugeordnet werden darf und daß obendrein diese Dokumentationseinheit das gesamte Dokument ist, gleichgültig wie reichhaltig und umfangreich es auch sein mag. Dies ist eine allenfalls früher zu rechtfertigende Konzession an die kargen Möglichkeiten der damals ausschließlich verfügbaren Speichermittel Steilkartei und Register. Aus einem solchen dürftigen Klassifikationsbegriff folgt für den Verfasser nur allzu logisch, daß die Facettenklassifikation dann gar keine Klassifikation ist, sondern lediglich eine Arbeitsweise mit mehreren Registern. Widersprüche tauchen dann auch zu der Frage auf, ob eine Facettenklassifikation nun eigentlich ein begriffskombinatorisches Prinzip ist oder nicht. Die Einteilung von Fachbegriffen nach einem sorgfältig ausgewählten Einteilungsgesichtspunkt fällt dann ebenfalls nicht mehr in das Gebiet der Klassifikation und bleibt fast gänzlich unerörtert. Es ist schade, daß gerade für diese spärliche und ineffektive Variante der Terminus „Klassifikation“ reserviert wird, wodurch der Klassifikationsbegriff schlechthin beim Leser und Lernenden in Mißkredit gerät. Es hilft auch nicht mehr viel, wenn diese Variante an einigen späteren Stellen als „Einfachklassifikation“ bezeichnet wird. Für eine spätere Auflage würde man sich größere Konsequenz in der Verwendung dieses verdeutlichenden Ausdruckes wünschen.

Zwar vermeidet der Verfasser ausdrückliche Hinweise auf die Schwächen der herkömmlichen Theorie. Daß diese aber so deutlich werden, ist nicht zuletzt der großen Übersichtlichkeit zu verdanken, in welcher er den Stoff gegliedert hat. So ist schon das Inhaltsverzeichnis durch eine graphische Darstellung ergänzt, in der die Relationen der einzelnen Kapitel zueinander wiedergegeben sind. Diese Kapitel sind hier in der Reihenfolge angeordnet, wie die in ihnen beschriebenen Schritte zeitlich nacheinander im Fluß der Informationsverarbeitung durchlaufen werden, angefangen von der Beschaffung der Dokumentationseinheiten bis hin zu ihrer Verwertung durch den Informationssuchenden.

Vorbildlich ist auch das Register, durch das der Stoff anscheinend lückenlos erfaßt erschlossen ist. Jedes Stichwort trägt als Ergänzung Hinweise, in welchem Zusammenhang es bei den betreffenden Textpassagen auftritt. Der Wert des Registers wird noch dadurch gesteigert, daß sich im Text selbst ein jedes Registerstichwort durch Kursivdruck zu erkennen gibt, ein einfaches und doch nur so selten anzutreffendes Hilfsmittel für einen Leser, der Passagen sucht, in denen ein bestimmtes Thema – vielleicht unter einem anderen Gesichtspunkt

– sonst noch abgehandelt ist. – Das Buch ist flüssig geschrieben, außerordentlich klar und übersichtlich gedruckt und praktisch frei von Druckfehlern. Es ist dem Lernenden, aber auch dem Praktiker zu empfehlen, denn es gibt viele wertvolle Ratschläge und Einblicke.

Robert Fugmann

Dr. R. Fugmann, Wiss. Dok., Hoechst AG
Postfach 800320, D-6230 Frankfurt 80

DOBBENER, Reinhard: **Grundlagen der Numerischen Klassifikation anhand gemischter Merkmale.** (Foundations of numerical taxonomy based on mixed variables) Göttingen: Vandenhoeck und Ruprecht, 1983. (Studien zur angewandten Wirtschaftsforschung und Statistik aus dem Institut für Statistik und Ökonometrie der Universität Hamburg; Heft 15) 143 p. ISBN 3-525-11286-6, DM 49,–

The present book is neither a textbook on numerical classification nor a treatise on a specific theory of numerical classification. One gets rather the impression that the author has tried to combine several approaches and results in one paper.

In the introduction several kinds of scales of measurement are introduced and discussed which are later on assumed for the variables, by which objects are to be classified. It seems that the author knows only in part the literature on measurement theory. In particular the example on p.16 can lead to confusion. However, this lack has no effect on the approaches to classification discussed later on.

The second chapter deals with invariance properties of metrics and classification criteria for interval and ratio scales. To this aim on the one hand those metrics are characterized which are invariant with respect to the transformations which are admissible with respect to interval and ratio scales. On the other hand an invariance property is defined for classification criteria in case of a fixed number of classes, and it is proved that the common variance and determinant criteria can be characterized by means of maximum likelihood by such properties. This approach is important and interesting for a theoretical evaluation and perhaps also for the construction of classification procedures. Unfortunately several proofs (p.25, 39, 47, 50) are not given in the book but only in an unpublished discussion paper of the author.

In the third chapter classification procedures are proposed for ordinal scaled variables and for the case of the simultaneous occurrence of nominal, ordinal, interval, and ratio scales. There is an immense need for such procedures in applications. The central point is the definition of an empirical variance criterion S for ordinal variables with a finite number of possible values. This is based on a representation of an ordinal variable with p distinct values by $p-1$ binary variables. An entropy criterion is calculated with respect to the cumulated relative frequencies of the ordered binary variables. It is proved that this criterion has quite a few interesting properties which illustrate its great importance in the context of classification procedures. Outside this context it would

be certainly helpful to consider first of all the properties of the corresponding population parameter instead of its empirical estimator. For data with variables measured on different scales it is proposed to classify interval and ratio scales in such a way that ordinal scales with a finite number of values result. In this way the problem is reduced to the evaluation of ordinal scales by means of the new entropy criterion and nominal scales by means of the usual entropy criterion. The proposal (p.98) to use normed entropy criteria in order to assign in this way equal weights to the different variables within the common variance is a bit problematic, at least in applications, since such artificial weights might lead to classifications which are difficult to interpret. But the given approach can easily be modified by introducing subjective and other weights. In particular, for ordinal scales properties of invariance and monotony are considered for a certain criterion and procedures for iterative and hierarchical classification based on this criterion are proposed.

In the fourth chapter statistical tests are discussed by which it should be tested whether a natural class structure is inherent to the data. On p.118a procedure is described which most probably will not lead to meaningful results. It is proposed to simulate data sets for which not only the vectors of measurement for each object but also the components of these vectors are independent. But it is self-evident that for concrete problems one should assume a certain dependence of the variables even under the null hypothesis of only one homogeneous class. This assumption is made in most known clustering procedures. Using the described approach in practice will have the effect that the probability of a false rejection of the null hypothesis is not controlled.

With respect to the procedure based on random permutations of the measurement vectors which is described on p.120 it seems that the author is not aware of the theory of randomization or permutation tests, respectively, originating from Fisher and Pitman, though he uses their fundamental idea. This can be concluded also from the very clumsy and time-consuming procedure on p.120 for generating random permutations. The application of such tests for evaluating classifications is by no means new, too (cf. e.g. Hubert et al., *Evaluation Review* 6(1982)p.505–520).

In contrast to the statement on p.122 it is possible in principle to determine the conditional distributions of the corresponding statistics by calculating for all possible permutations the value of the test statistic and assuming a discrete uniform distribution under the null hypothesis. In this way exact tests are possible without the need for a simulation study. With respect to today's computers such tests, of course, can be performed only for rather small samples and in general one will fall back on simulation studies. However, for a reliable evaluation some thousands of replications are necessary, since on the basis of only 6 permutations (cf. section 4.4) no justified conclusions are possible with respect to the structure and number of classes.

Altogether one should regard the present book rather as a kind of research report and not as a fundamental