

REPORTS AND COMMUNICATIONS

DIN 31 623 oder die Problematik des genormten Indexierens (Concerning the problematics of indexing standardization)

by Dr. Robert Fugmann

(Ed. Note: In November 1978 the German Institute for Standardization (DIN) issued a Draft Standard named DIN 31 623 "Indexierung zur inhaltlichen Erschließung von Dokumenten" (Indexing for subject access to documents) in two parts: Pt. 1 is concerned with relevant indexing concepts and criteria of indexing, Pt. 2 tries to submit rules for coordinate indexing with descriptors and criteria for concept delimitation by descriptors. For some time a number of German classification experts have been opposing the intentions and solutions contained in this draft standard. They were not listened to. We should like to give one of them a chance to argue here for a greater awareness of the deficiencies in such standards which may lead to a continual decline in the functioning of information systems. It seems timely to use a public medium for the necessary corrective statements in order to counteract an inconsiderate acceptance of this draft as a standard. We beg our English readers' pardon that the following text is in the same language as the document questioned.)

A. Einführung

A.1 Beim Deutschen Institut für Normung ist eine Arbeitsgruppe seit längerer Zeit damit beschäftigt, eine Norm für das Indexieren von Dokumenten aufzustellen. Insbesondere soll das Indexieren von wissenschaftlicher und technischer Literatur einer solchen Norm unterliegen. Auf die Problematik einer solchen Zielsetzung angesprochen, hat das DIN den wissenschaftlichen Charakter, den die Indexierungsarbeit zumindest in ihren höher entwickelten Varianten besitzt, in Abrede gestellt und die Normierbarkeit des Indexierens konstatiert.

A.2 Durch das Indexieren soll aber die Grundlage für das treffsichere und wirtschaftliche Wiederauffinden der jeweils gewünschten Fachinformation gelegt werden. Dies setzt aufseiten des Indexers das volle Verständnis der laufend erscheinenden Fachliteratur voraus, sowie auch meistens den fortgesetzten, sachverständigen Dialog mit den Wissenschaftlern und Technikern als Informationssuchenden. Auch muß der ständige Wandel im Informationsbedarf seinen Niederschlag in einer fortgesetzt angepaßten Indexierungsweise finden. So stellt das Indexieren einen Zweig der fachwissenschaftlichen Arbeit selbst dar, und obendrein ist es auch eine grundlegende informationswissenschaftliche Tätigkeit. Nicht umsonst sind in weiten und durchaus kostenbewußten Kreisen in Wirtschaft und Technik Wissenschaftler laufend mit Indexierungsarbeiten beschäftigt.

A.3 Das DIN entwickelt aber auch auf anderen Gebieten, deren wissenschaftlicher Charakter nicht bestritten wird, ähnlich verfehlte Aktivitäten, und hieran ist vom Verfasser bereits an anderer Stelle Kritik geübt worden¹. Dieses Vordringen von Normen führt, wenn ihm nicht entgegengetreten wird, auf einen Weg, an dessen Ende vielleicht eines Tages eine Norm zur Behandlung von Krankheiten beim Menschen, zur landwirtschaftlichen Nutzung von Bodenfläche oder zur Herstellung von Arzneimitteln liegt. Dies wäre nur die Ausdehnung der derzeitigen verfehlten Normierungsversuche im wissenschaftlichen Bereich auf andere Gebiete.

A.4 Die Fachleute auf informationswissenschaftlichem Gebiet dürfen sich nicht damit beruhigen, daß solche Normen für die wissenschaftliche Arbeit nicht maßgebend sein können und demzufolge ohnehin nicht angewendet würden und daher auch keinen Schaden anrichten könnten. Aber durch die Verbreitung solcher Normen, wie es z.B. gelegentlich auch schon im Unterricht geschieht, würde in der Öffentlichkeit der Eindruck entstehen, daß sich der Indexer vermeintlich nur nach einer Norm zu richten braucht. Auch würde das verbreitet anzutreffende Mißverständnis vom leicht erlernbaren Indexieren weitergenährt werden, von einem Indexieren, das ohne Nachteile auf allen Gebieten und für fast jeden Zweck auch Hilfskräften oder gänzlich der Maschine überlassen werden könne. Dieselbe Öffentlichkeit, welche heute – wenn auch unbewußt – solche abwegigen Normierungs- und Reglementierungsversuche der wissenschaftlichen Arbeit toleriert und gar fördert, könnte eines Tages auch auf der Durchsetzung solcher Normen bestehen. Wenn dann – wenn auch nur allmählich und daher nicht so spektakulär – die Funktionstüchtigkeit der betroffenen Informationsbereitstellungssysteme immer mehr verloren geht, dann wird man sich in derselben Öffentlichkeit lange der Einsicht in die tieferen Ursachen verschließen, welche im Festschreiben der durch die Normen propagierten Irrtümer und Mißverständnisse liegt.

A.5 Aber nicht nur wegen ihrer verfehlten Zielsetzung muß dem Normentwurf für das Indexieren widersprochen werden. Auch in den Details des Entwurfs spiegeln sich Fehler wider, wie sie sich leicht in die informationswissenschaftliche Arbeit einschleichen können. Bleiben diese Ausführungen unwidersprochen, so können sie den weniger Erfahrenen leicht auf Irrwege führen. Es ist für diese Irrwege charakteristisch, daß sie oftmals erst sehr spät als solche erkannt werden und daß sie dann nur noch sehr schwer und mit großem Aufwand und Zeitverlust wieder verlassen werden können. Oft ist sogar ein völliger Neubeginn unter weitgehendem Verzicht auf die Früchte der auf diesem Irrweg geleisteten Arbeit notwendig.

A.6 Im Rahmen dieser Kritik begnügen wir uns damit, zwei beliebig herausgegriffene Details aus den Teilen I und II des Normentwurfs zu unterscheiden und zu fragen

1. wie die Eigenarten von verbindlichem und freiem Indexieren dargestellt und
2. wie die Empfehlungen zur Abgrenzung der Vokabularbegriffe formuliert wurden.

A.7 Zu fast jedem anderen Detail der Norm läßt sich eine ähnlich gravierende und ausgedehnte Kritik vortragen².

B. Verbindliches vs. freies Indexieren

B.1 Aus dem Normentwurf geht nicht klar hervor, was dort unter Indexieren mit freiem und verbindlichem Vokabular verstanden wird, da die entsprechenden Definitionen und Erläuterungen fehlen, bzw. im Widerspruch zueinander stehen. So wird für das Indexieren mit freiem Vokabular nach I, Seite 3; 5.7.1 die Benutzung eines vorgegebenen Vokabulars *offen* gelassen. In II, Seite 3; 4.1.3 wird auch beim freien Indexieren die Verwendung eines Vokabulars „als Indexierungshilfsmittel (Nachschlagewerk)“ *verlangt*, wenn auch nicht in dem Sinne verbindlich, daß die jeweils bestpassenden Deskriptoren in ihm aufgesucht werden müssen. Nach II, Seite 2; 3. sind auch beim freien Indexieren die Regeln II, 4. und 5. zu beachten, von denen ein Teil auch für das verbindliche Indexieren gilt, nämlich 4.1., 4.2.1; 4.2.2. Noch andere Passagen erfordern auch beim Indexieren mit einem freien Vokabular den *verbindlichen Einsatz eines vorgegebenen Vokabulars*, mag dort auch die Notwendigkeit des verbindlichen Einsatzes nicht immer ausdrücklich genannt sein. Dies gilt für die „Begriffszusammenführung“ (II, Seite 17; 5.5.), für die Wahl von negierenden Deskriptoren (II, Seite 7; Beispiel 15), für die Erkennung von entbehrlichen, weil hierarchisch verwandten Deskriptoren (II, Seite 4; Beispiel 4), sowie auch für die gemäß den Empfehlungen durchzuführende „Begriffszerteilung“ (II, Seite 14; 5.4.3.2.1). Im Widerspruch zu dieser *vorgeschriebenen Orientierung nach dem Vokabular* wird an anderer Stelle für das freie Indexieren die *Orientierung nach dem Text des Dokumentes* verlangt (z.B. II, Seite 4; 4.2.3.1: „Die Deskriptoren sind – mit Ausnahme offenkundiger Fehler und soweit sie nicht am Rande des Dokumentationsgebietes liegen – in möglichst enger *Anlehnung an Titel und Text des Dokumentes* zu wählen“*). – Wenn für das freie Indexieren zwei genau entgegengesetzte Maßnahmen empfohlen werden, so kann man hierin keine brauchbare Anleitung zum praktischen Handeln erblicken.

B.2 Unter den Vorteilen des freien Indexierens (I, Seite 3; 5.7.1) wird in der Norm hervorgehoben, daß auf diese Weise eine „*schnelle und spezifische Inhaltskennzeichnung*“ erreicht werden kann*. Es kommt aber beim Indexieren³ weniger auf die *Inhaltskennzeichnung*⁴ als vielmehr darauf an, daß man später einmal voraussehen oder rekonstruieren kann, wie dieser Inhalt sprachlich ausgedrückt worden sein könnte. Genau diese sprachliche Ausdrucksweisen sind es nämlich, die man beim Formulieren einer Fragestellung kennen muß. Ein Hauptgrund dafür, weshalb man überhaupt indexieren muß, liegt ja darin, daß diese Ausdrucksweise im Originaltext nicht voraussehbar oder rekonstruierbar ist.

B.3 Das Kriterium der Voraussehbarkeit oder Rekonstruierbarkeit, das für das Indexieren und Wiederfinden so wichtig ist, wird aber im Normentwurf überhaupt nicht erwähnt. Stattdessen wird die Reproduzierbarkeit⁵ des Indexierens in den Vordergrund gestellt. Es ist aber leicht zu erkennen (vgl. Nachr. Dok. 29 (1978), S. 122 und Anh. 1.), daß die Forderung der Reproduzierbarkeit für das Indexieren nicht ausreicht und daß diese Lücke auch nicht durch die Kriterien der Wiedergabetreue, Eindeutigkeit, Prägnanz und Gebräuchlichkeit⁶

der Deskriptoren geschlossen wird. Was die Gebräuchlichkeit der Deskriptoren anbetrifft, so ist auch dies wieder eine Forderung, die für das Wiederfinden belanglos ist und lediglich im Interesse desjenigen liegt, der die Deskriptoren *lesen* will, um sich über den Inhalt eines Dokumentes zu informieren. Andernfalls müßten die (völlig ungebrauchlichen) kunstsprachlichen Deskriptoren, z.B. Notationen aus einer speziellen Klassifikation, für das Indexieren und Wiederfinden ungeeignet sein.

B.4 Was die „*spezifische*“ Inhaltskennzeichnung anbetrifft, so trifft es nicht zu, daß sie bevorzugt durch freies Indexieren erreichbar ist. Dies wird bei der Betrachtung des Beispiels 4 in II, Seite 4 deutlich:

„Durch mitgeschleppte Fremdkörperpartikeln kann es in kupfernen Wasserleitungsrohren zu lokal scharf begrenzten Zerstörungszonen kommen. Es wird der Einbau von Filtern empfohlen, die vorzugsweise aus feinen Polyamidgeweben bestehen“

B.5 Enthält ein Indexsprachenwortschatz für das Gebiet von Korrosion und Korrosionsschutz die Deskriptoren

Wasserfilter	Kupferkorrosion
Gewebefilter	Wasserleitungsrohre
Polyamidgewebe	Kupferrohre
	Lochfraß ⁷ -Verhütung

so wird man *beim verbindlichen Indexieren diese Deskriptoren auch verwenden*, weil mit ihnen der Inhalt dieses Dokumentes sehr spezifisch wiedergegeben werden kann und weil das verbindliche Indexieren die Benutzung der jeweils bestpassenden Deskriptoren aus dem Vokabular der Indexsprache verlangt.

B.6. Beim freien Indexieren orientiert man sich hingegen nicht nach dem Wortschatz, sondern nach dem Text der Vorlage (wie es in II, Seite 4; 4.2.3.1 unter „Auswahl freier Deskriptoren“ empfohlen wird). Dann gelangt man zu den *viel weniger spezifischen* Deskriptoren (wie sie auch von der Norm für dieses Dokument empfohlen werden (vgl. II, Seite 4; 4.2.2, Beispiel 4):

Kupfer	Filter
Wasserleitungsrohr	Polyamid
lokale Zerstörung ⁸	Gewebe

B.7 Es trifft also nicht zu, daß die Indexierungsspezifität beim freien Indexieren, d.h. ohne Orientierung an einem vorgegebenen Vokabular, grundsätzlich höher ist als beim verbindlichen Indexieren. Eine wichtige Ursache hierfür besteht darin, daß beim freien Indexieren nicht mit Zuverlässigkeit die später noch zu erörternde „Begriffszusammenführung“ eintritt (vgl. Abschnitt E). Deswegen kann das freie Indexieren leicht zum Verlust an einschlägiger Information führen. Die oben genannten spezifischen Deskriptoren „Wasserfilter“, „Gewebefilter“ usw. werden nämlich vom Wortschatz zumindest *für das Formulieren einer Fragestellung* zum gleichen Thema angeboten. Ein solches Angebot zu unterbreiten ist sogar die eigentliche Aufgabe des Wortschatzes. Der Informationssuchende wird dann von diesen spezifischen Deskriptoren im Interesse einer ausreichend spezifischen Fragestellung auch Gebrauch machen, im Vertrauen darauf, daß sie vom Indexer ebenfalls verwendet worden sind. Dann können aber die einschlägigen, jedoch (normgemäß) unspezifisch indexierten Dokumente nicht aufgefunden werden. Wenn man versucht, durch eine Verallgemeinerung der Fragestellung Abhilfe zu schaffen, so kommt zu diesem Verlust noch der Ballast hinzu (vgl. Anhang Abschnitte 2. und 3.).

* Hervorhebungen von Verfasser.

B.8 Es trifft auch nicht zu, daß das freie Indexieren grundsätzlich *schneller* als verbindliches Indexieren ist. Wenn man nämlich als Indexer gewissenhaft ist und sich bemüht, auch in Abweichung vom Text der Vorlage möglichst spezifische Deskriptoren zu finden, wie es von der Norm ebenfalls für das freie Indexieren empfohlen wird (II, Seite 4; 4.2.2.)⁹, dann kann der Zeitaufwand für die günstigste Deskriptorenwahl sehr groß werden. Erschwerend wirkt hierbei, daß der Indexer über kein Kriterium dafür verfügt, wann er mit dem Nachdenken über noch spezifischere, noch prägnantere Deskriptoren aufhören kann¹⁰.

B.9 Beim verbindlichen Indexieren mit einem vorgegebenen Vokabular hingegen entfällt dieses Kopfzerbrechen, und zugleich hat der Indexer auch die Gewißheit, daß auch jemand anders als er selbst mit den Deskriptoren aus dem Vokabular (oder mit einer Kombination von ihnen) das indexierte Dokument leicht und sicher wiederfinden wird. Es handelt sich dann nämlich nicht um Deskriptoren, die ihm eingefallen sind (und die vielleicht einem anderen Indexer nicht eingefallen wären). Vielmehr ist durch die Anlehnung an das Vokabular die Deskriptorenwahl rekonstruierbar geworden. Dieses wichtige Kriterium der Rekonstruierbarkeit oder Voraussehbarkeit der Deskriptorenwahl bleibt jedoch in dem Normentwurf gänzlich außer Betracht (vgl. auch Anhang, Abschn. 1.).

B.10 Wenn beim verbindlichen Indexieren die „Umsetzung der im Dokument enthaltenen Begriffe in eine vorgegebene Dokumentationssprache“ als Last empfunden wird (I, Seite 3; 5.7.1., 2. Abs.), dann sollte der Indexer lieber inhaltkomprimierende Zusammenfassungen oder Abstracts schreiben. Arbeitet man nach den Empfehlungen der Norm, so muß man auch beim freien Indexieren obendrein das Vokabular so oft und so intensiv zu Rate ziehen, daß dies bereits der „Umsetzung in eine vorgegebene Dokumentationssprache“ fast gleich kommt (vgl. B.1)

B.11 In der Normvorlage werden aber auch einige Nachteile des freien Indexierens erwähnt (I, Seite 4; 5.7.1), und es werden zwei Alternativen zu ihrer Überwindung angeboten. Die eine von ihnen empfiehlt beim Frageformulieren das „Nachschlagen im Vokabular, Auffinden von Formulierungsalternativen“, wobei auch richtig erwähnt wird, daß dies eventuell nicht zum Ziele führt.

B.12 Bei der zweiten Gegenmaßnahme wird diese Einschränkung nicht gemacht, und es entsteht der Eindruck, daß mit ihr die Nachteile des freien Indexierens überwindbar sind. Es wird empfohlen,

„daß die indexierten Deskriptoren anschließend an die Indexierung kontrolliert und (soweit nicht schon enthalten) in ein erweiterungsfähiges Vokabular eingeordnet und im Speicher durch die entsprechenden Begriffsrepräsentanten des Vokabulars ersetzt werden“¹¹.

B.13 Es ist aber leicht zu erkennen, daß auch mit dieser Gegenmaßnahme nur ein Teilerfolg erzielbar ist. Hat nämlich der Indexer einmal die (viel zu unspezifischen) Deskriptoren „Filter“, „Gewebe“, „Kupfer“, „Polyamid“, „Wasserleitungsrohr“, „lokale Zerstörung“, die von der Norm empfohlen werden (vgl. II, Seite 4, Beispiel 4), vergeben, dann kann keinerlei nachträgliche „Kontrolle“ irgendwelcher Art noch zu dem besser treffenden, spezifischeren Deskriptoren aus dem Vokabular führen (etwa zu „Lochfraßverhütung“, „Gewebefilter“,

„Wasserfilter“ usw.). Dies aber sind die Deskriptoren, die der Informationssuchende bei einem einschlägigen Dokument erwartet, wenn diese Deskriptoren im Vokabular vorkommen und zum Formulieren einer thematisch analogen Fragestellung geradezu angeboten werden. Daß nämlich das Kupfer der Korrosion unterliegt (und nicht etwa der Werkstoff Polyamid oder ein anderes im Dokument vielleicht genanntes Material), daß es sich auch nicht lediglich um eine Zerstörung irgendwelcher Art handelt, daß weiterhin das Filter aus Gewebe besteht, und zwar aus Polyamid usw., kann nicht mehr aus den unspezifischen, zusammenhanglos aufgezählten Deskriptoren „Filter“, „Gewebe“, „Kupfer“ usw. alleine beurteilt werden. *Vielmehr wäre die nochmalige Einblicknahme in das bereits indexierte Dokument erforderlich*, denn nur aus diesem sind die Zusammenhänge ersichtlich, die durch die spezifischen Deskriptoren ausgedrückt werden sollen. Wenn man also „die indexierten Deskriptoren anschließend an die Indexierung kontrolliert“ (I, Seite 4; 5.7.1), dann ist in diesem Arbeitsstadium nur noch eine evtl. „Begriffszerlegung“ möglich, jedoch nicht die ebenso wichtige „Begriffszusammenführung“ (II, Seite 17; 5.5 vgl. E und Anhang, Abschnitt 2.3).

B.14 Die Nachteile des freien Indexierens können also zuverlässig nur durch die Verwendung eines vorgegebenen, zur verbindlichen Anwendung vorgeschriebenen Vokabulars überwunden werden. Diese Schlußfolgerung hätte in aller Deutlichkeit im Normentwurf ausgedrückt werden müssen, und *die gesamte Prozedur der Deskriptorenwahl hätte von Grund auf anders organisiert werden müssen, als es in der Norm geschehen ist*. Am Anfang der Indexierungsarbeit hätte nämlich nicht die Wahl von – ohnehin nur provisorisch festgelegten (vgl. E.5) – Deskriptoren in Anlehnung an den Text des Dokumentes stehen dürfen, sondern vielmehr das verbindliche Indexieren des Dokumententextes nach einem vorgegebenen Vokabular (vgl. B13, E3, E5).

B.15 In der Praxis muß aber nicht immer auf hochgradig verlustarmes Recherchieren Wert gelegt werden. Auch hat man es dort oft mit Individualbegriffen und schlecht definierten Termini zu tun. Deshalb wird dort verschiedentlich das *wirklich freie Indexieren* ohne die Orientierung nach einem vorgegebenen Vokabular vorgezogen. Eine solche Arbeitsweise wird aber von den Empfehlungen der Norm in reiner Form überhaupt nicht beschrieben.

B.16 Auch die Arbeitsweise des *wirklich verbindlichen Indexierens* wird in der Norm nicht dargestellt, offenbar weil man das Aufsuchen der jeweils bestpassenden Deskriptoren im Vokabular als einen „mechanischen Vorgang“ (II, S. 3; 4.1.3.) ansieht und ihn keiner besonderen Erörterung für würdig erachtet. Man hat in der Norm nicht erkannt, wie problematisch gerade dieser Suchprozeß werden kann. Ein Wortschatz nämlich, wie er sich aufgrund der Norm, insbesondere aufgrund ihres Regelwerkes zur Begriffsabgrenzung entwickelt, wird auf vielen Gebieten bald nicht mehr verbindlich handhabbar sein.

B.17 Dadurch, daß die Norm die Stärken und Schwächen von verbindlichem und freiem Indexieren nicht treffend erkennt (wovon hier nur zwei Beispiele von vielen anderen aufgeführt sind), gelangt sie auch nicht zu einer wirklich sinnvollen *Kombination* und Aufgaben-

teilung *beider Indexierungsarten*, wie sie in der Praxis heute verbreitet anzutreffen ist, einer Aufgabenteilung nämlich, in der jede der beiden Indexierungsarten genau in dem Bereich eingesetzt wird und auf den Bereich beschränkt bleibt, in dem sie am wirkungsvollsten arbeitet. Stattdessen empfiehlt die Norm ein Konglomerat aus beiden Indexierungsarten mit widersprüchlichen Anweisungen. Die Befolgung der einen Regel hat zwangsläufig den Verstoß gegen eine andere, gleichrangige im Gefolge. Als Anleitung zum praktischen Handeln ist ein solches Regelwerk ungeeignet.

C. Die Begriffsabgrenzung

C.1 Als zweites Beispiel für ein unzulänglich ausgeführtes Detail im Normenentwurf betrachten wir das dort beschriebene Verfahren zur Begriffsabgrenzung. Wenn man einen Indexsprachen-Wortschatz verbindlich einsetzen will, wie es von dem Normenentwurf durch verschiedene Regelungen auch für das freie Indexieren verlangt wird (siehe Abschnitt B.1), so muß dieses Vokabular bezüglich seines Umfangs und seiner Struktur eine wesentliche Voraussetzung erfüllen. Es darf nicht zu viele und allzu komplexe¹² Begriffe in Gestalt der betreffenden präkombinierenden (oder präkoordinierten) Deskriptoren umfassen, und der Zustrom an neuen Deskriptoren zum Vokabular muß durch entsprechende Regelungen gedrosselt sein.

C.2 Ob sich ein Deskriptor für die Aufnahme in den Wortschatz gut eignet, hängt in erster Linie vom Grad seiner Komplexität ab (vgl. Anhang, Abschnitt 4.). Diese Komplexität spiegelt sich aber nicht darin wider, mit wievielen Wörtern ein Begriff ausgedrückt (oder ausdrückbar) ist, wieviele Wortstämme oder Lexeme in einer solchen Ausdrucksweise vorkommen, oder ob Adjektive („korrosiv“), Substantive („Korrosion“), Verbformen („korrodieren“) usw. in ihr auftreten. Für ein Urteil über die Komplexität eines Begriffes ist vielmehr allein ausschlaggebend, welche und wie viele begriffliche Komponenten bestimmter Art er enthält (vgl. auch II, S. 9; 5.2.b). Erweist sich ein wesentlicher Begriff als unerwünscht komplex, so sollte er nicht in den Wortschatz aufgenommen werden. Stattdessen sollte er durch die in ihm enthaltenen begrifflichen Komponenten dargestellt werden, *sowie möglichst auch durch die Wiedergabe des engen Zusammenhanges, der zwischen den begrifflichen Komponenten besteht*. Ohne eine solche Syntax wäre die Wiedergabetreue der Indexierung unbefriedigend. Es sei nur am Rande erwähnt, daß das letztgenannte Erfordernis in der Norm völlig außer Betracht geblieben ist (vgl. Anhang Abschnitt 5.). Dieses ist die Ursache für eine ganze Serie von weiteren Mängeln der Norm, auf die hier jedoch nicht näher eingegangen werden soll (Vgl. E.2, E.7, E.8).

C.3 Wenn ein Begriff, der natursprachlich durch ein einziges Wort (oder durch eine einzige Wortfügung bestimmter Art) ausgedrückt ist, durch mehr als einen einzigen Deskriptor wiedergegeben wird, so wird dies im Normenentwurf als „Begriffszerlegung“ bezeichnet. Wählt man umgekehrt anstelle mehrerer Deskriptoren aus dem Vokabular einen präkoordinierten Deskriptor für den entsprechenden komplexen Begriff, so wird dies als „Begriffszusammenführung“ bezeichnet¹³.

D. „Begriffszerlegung“

D.1 Um das Regelwerk der Norm zur „Begriffszerlegung“ zu durchschauen, vergegenwärtigen wir uns den eigenartigen Widersinn des nachfolgenden hypothetischen Textes:

Teil A (richtig):

„Für die Zuteilung von speziellen Aufgaben sind Ausbildung und Arbeitsstil eines Mitarbeiters maßgebend. Seine *formal-äußerlichen Merkmale dürfen hierbei keine Rolle spielen*.“

Teil B (falsch):

„Damit bei der Zuteilung von speziellen Aufgaben nicht jeder Mitarbeiter auf seine Ausbildung und seinen Arbeitsstil geprüft werden muß, ist es sinnvoll, eine *formal-äußerliche Untersuchung vorzuschalten*, durch welche solche Mitarbeiter erkannt und abgetrennt werden können, die für die Zuteilung solcher Aufgaben nicht infrage kommen.“

und

„Bei der Überprüfung, ob einem Mitarbeiter *überhaupt* spezielle Aufgaben übertragen werden können, ist es sinnvoll, als *formalen Indikator* die Anzahl der Vokale in seinem Familiennamen zu betrachten.“

D.2 Der entwaffnende Widersinn dieses Textes liegt darin, daß den formal-äußerlichen Merkmalen eines Menschen eben doch – und im Widerspruch zu dem gleichzeitig konstatierten Prinzip – ein Einfluß auf die Aufgabenteilung eingeräumt wird. Es handelt sich sogar um einen ganz wesentlichen Einfluß, insofern nämlich, als es von formalen Merkmalen entscheidend abhängig gemacht wird, ob ein Mitarbeiter *überhaupt* für die Zuweisung spezieller Aufgaben in Betracht gezogen werden soll.

D.3 Die gleiche Art von Widersinnigkeit ist im Regelwerk zur „Begriffszerlegung“ in DIN 31623 zu verzeichnen. Der größeren Deutlichkeit halber seien die dort gegebenen Regeln ebenfalls in zwei Teilen zusammengetragen und gegenübergestellt:

Teil A* (richtig):

Bei der *Begriffsanalyse* und der ggf. notwendigen *Begriffszerlegung dürfen formale Gesichtspunkte keine Rolle spielen*“ (II, S. 10; 5.4.).

Teil B* (falsch):

„Damit bei der Indexierung nicht jeder Fachterminus einer *Begriffsanalyse* unterzogen werden muß, ist es sinnvoll, eine *sprachlich-formale Analyse* vorzuschalten, durch die Benennungen erkannt und abgetrennt werden, die für die Zerlegung nicht infrage kommen (II, S. 10; 5.3.).“

und im Kapitel zur Begriffsanalyse (II, S. 13; 5.4.2.):

„Bei der Überprüfung, ob eine Zerlegung möglich oder notwendig ist, ist es sinnvoll, folgende drei Kategorien von *formalen Indikatoren* zu betrachten . . . :

- Anzahl der in der Benennung enthaltenen Lexeme
- Sprachform der Deskriptoren

...“

D.4 Es ist den Verfassern der Norm offenbar entgangen, in einer wie weitgehenden Weise sie die ganze „Begriffszerlegung“ nun doch von formal-sprachlichen Gesichtspunkten abhängig gemacht haben. Es sind nämlich nun die formal-sprachlichen Kriterien, von denen es einzig und allein abhängt, ob eine „Begriffszerlegung“ überhaupt in Betracht gezogen werden soll. Alle hochgradig komplexen Begriffe, die eine Begriffsanalyse dringend nötig hätten, entziehen sich nun der Begriffsanalyse, sofern sie sich nur in einer lexemarmen, bzw. wortstammarmen Ausdrucksweise formulieren lassen, und sofern diese Ausdrucksweise dem Autor oder dem Indexer eingefallen ist.

D.5 Dieser Widerspruch zwischen einzelnen Teilen des Regelwerkes tritt offen oder in versteckter Form auch

noch an anderen Stellen des Regelwerkes zutage, z.B.

Textart A (richtig)*:

„Soll beim Indexieren eine Zerlegung vorgenommen werden, so müssen stets *Begriffe* in weniger komplexe Einzelbegriffe zerlegt werden, *nicht Begriffsbenennungen* in ihre sprachlichen Elemente (II, S. 10; 5.2.)“.

Textart B (falsch):

„Bei der Erwägung, ob ein *Begriff* zerlegt werden muß, ist von der *Benennung* auszugehen, die den gemeinten Begriff möglichst treffend . . . wiedergibt“ (II, S. 11; 5.4.1.).

und

„Ob eine *Benennung begriffsanalytisch* geprüft werden muß, wird nach der Zahl der in ihr enthaltenen Lexeme entschieden“ (II, S. 10; Beispiel 34), wobei die Bestimmung der Lexemzahl auch nach II, S. 10; 5.3. eine „sprachlich-formale“ Operation ist.

D.6 Wieder wird zunächst richtig die Irrelevanz einer sprachlich formalen Analyse für den Prozeß der „Begriffszerlegung“ konstatiert. Zugleich wird jedoch (und in einem Fall sogar in ein und demselben Absatz (II, S. 10; 5.4.) der sprachlich formalen Analyse sogar die Entscheidung der Grundfrage überlassen, ob in einem fraglichen Fall die Begriffsanalyse überhaupt durchgeführt werden soll.

D.7 Diese sprachlich-formale Analyse besteht in der Untersuchung, wie viele Wortstämme in einer (nach den speziellen Regeln der Norm und auch nur vorläufig festzulegenden) natursprachlichen Ausdrucksweise enthalten sind (vgl. II, S. 12; 5.4.1.). Hierzu „wird der Begriff des Lexems eingeführt“ (II, S. 10; 5.3.)¹⁴. „Eingeführt“ ist hierbei aber in dem Sinne zu verstehen, daß der Lexembegriff in der Indexierungspraxis in der Tat ungewöhnlich ist. Der Grund für den Verzicht auf die Lexemanalyse beim Indexieren liegt aber darin, daß es zwischen der natursprachlichen Ausdrucksweise für einen Begriff einerseits und seiner Komplexität auf der anderen Seite keinerlei verwertbare Beziehung gibt. Legt man nämlich den Lexembegriff zugrunde, dann würde die „Begriffszerlegung“ in unerlaubter Weise von zufälligen, peripheren Merkmalen abhängig gemacht (vgl. Anhang, Abschnitte I. und 4.).

D.8 Unklar bleibt die Rolle einer nochmals anderen, ebenfalls nach rein sprachlich-formalen Kriterien beschriebenen Gruppe von Begriffen. Es sollen nämlich

„einfach formulierte, in der Fachsprache geläufige Begriffe („Lexikonbegriffe“)¹⁵ möglichst nicht zerlegt werden“ (II, S. 10; 5.2.),

und zwar, wie es in der Norm heißt,

„um die Indexierung nicht unnötig zu erschweren“¹⁶.

Jedoch heißt es schon im nächsten Satz:

„In welchen Fällen semantisch zerlegt werden soll, wird in Abschnitt 5.4. behandelt“.

D.9 Dieser Abschnitt beschreibt jedoch das in der Norm generell empfohlene Regelwerk zur (semantischen) Begriffsanalyse, so daß nun wieder unklar ist, ob überhaupt die „Lexikonbegriffe“ eine Sonderstellung einnehmen sollen.

D.10 Mit dem „Lexikonbegriff“ wird dem Indexer immerhin die Möglichkeit geboten (oder es wird ihm in undefinierter Weise sogar zur Auflage gemacht) wenigstens den abwegigsten Forderungen der Lexemanalyse auszuweichen. Eigentlich müßte sich der Indexer den Kopf darüber zerbrechen, wie viele (und welche!) Lexeme in Ausdrücken wie „Rotbarsch“, „rotbuntes Tieflandrinde“, „Roteisenstein“ (bedeutungsgleich mit Hämatit), „Rot-halstaucherweibchen“, „Würzburger Löwenbräu“ usw.

zu identifizieren sind. Er wird sich fortgesetzt fragen, welchen Sinn eine solche Maßnahme haben soll, die deswegen so fruchtlos ist, weil man aus der sprachlichen Formulierung für einen Begriff sowieso keinerlei Rückschlüsse auf seine Komplexität ziehen kann.

D.11 Ein Höhepunkt der Verwirrung wird durch den Satz erreicht (II, S. 10, 5.3)*:

„Ein Lexem verkörpert somit eine sprachlich selbständige begriffliche Einheit, die nach *morphologischen Methoden* nicht weiter in *semantische Komponenten* zerlegt werden kann“.

D.12 Aber wie kann man überhaupt versuchen, mit morphologischen Methoden zuverlässig zu *semantischen Komponenten* zu gelangen, wo doch die morphologische Zerlegung höchstens in Ausnahmefällen mit den Ergebnissen der semantischen Zerlegung übereinstimmt (vgl. II, S. 10; 5.2.)?

D.13 Wendet man das Regelwerk der Norm zur „Begriffszerlegung“ an, so gelangt man auf dem langen und widerspruchsvollen Weg der – zunächst nur vorläufigen – Deskriptorenwahl und der sprachlich-formalen Analyse (II, S. 2–14; Regeln 4. und 5.) zu einem sonderbaren Ergebnis: Viele hochgradige komplexe Begriffe haben sich der Begriffsanalyse entzogen, weil sie zufällig sprachlich lexemarm formuliert gewesen sind (z.B. als „Hämatitlagerstätte“ und nicht als „Roteisensteinlagerstätte“, als „Messingkorrosion“ und nicht als „Korrosion von Legierungen aus Kupfer und Zink“ im Verhältnis x : y usw.). Auf diese Weise gelangen auch viele hochgradig komplexe Begriffe ins Vokabular, was dem Sinn der Begriffsanalyse zuwiderläuft. Ein normgemäß aufgebauter Wortschatz wird daher bald allzu viele komplexe Begriffe enthalten und bald nicht mehr verbindlich anwendbar sein. Die gesamte komplizierte, auf 10 Seiten der Norm beschriebene Prozedur verfehlt also weitgehend den Zweck, für den sie eigentlich gedacht ist.

D.14 Aber auch die „Begriffszusammenführung“ kann nach der normgemäß durchgeführten Deskriptorenwahl und der formal-sprachlichen Analyse nicht mehr in geregelter Weise durchgeführt werden. Hierfür muß man nämlich vom Originaltext ausgehen, wie nachfolgend dargelegt wird.

E. Die „Begriffszusammenführung“

E.1 Bei einem jeden wesentlichen Fachausdruck muß man sich entscheiden, ob man den zugehörigen Begriff „zerlegen“ oder aber mit einem anderen, im Text ebenfalls ausgedrückten oder deutlich implizierten Begriff „zusammenführen“ soll (Vgl. Anhang, Abschn. 3). Daß beide Verfahren gleichrangig sind, wird zwar in II, S. 9; 5.1. betont, aber es wird keine Entscheidungshilfe gegeben, welche der beiden konträren Maßnahmen im Einzelfall zu wählen ist.

E.2 Im Falle des Beispiels 4 in II, S. 4; 4.2.2. (vgl. B.4) müßte man eigentlich konstatieren, daß der Sachverhalt „sprachlich in so stark zerlegter Form ausgedrückt“ ist, „daß eine Begriffszusammenführung zweckmäßig ist“ (II, S. 17; 5.5.) Vor allem sollten die wichtigen Relationen zwischen dem Kupfer und dem Korrosionsvorgang, die Funktion des Filters und des Polyamidgewebes (nämlich als Filtermaterial usw.) zum Ausdruck gebracht werden. Steht für diesen Zweck keine Indexsprachensyntax zur Verfügung, wovon die Norm ausgeht, so bleibt

nur die Bildung von präkoordinierten Deskriptoren, wie z.B. „Wasserfilter“, „Gewebefilter“, „Kupferkorrosion“ übrig, falls hohe Wiedergabetreue verlangt wird. Es widerspricht diese Begriffszusammenführung der Regel in II, S. 17; 5.5., daß diese Begriffszusammenführung unterblieben ist und daß man stattdessen die hochgradig zerlegte, aber viel zu ungenaue Indexierungsweise bevorzugt hat. (vgl. C. 2 und Anhang, Abschn. 5).

E.3 Es ist zwar richtig, daß eine „geregelte Begriffszusammenführung“ ein vorgegebenes verbindliches Vokabular voraussetzt (II, S. 17; 5.5.), aber man vermißt einen Hinweis darauf, an welcher Stelle des gesamten Prozesses der Begriffsabgrenzung dieses Vokabular zur Anwendung kommen muß. Die „Begriffszusammenführung“ muß nämlich vollzogen werden, *bevor* das gesamte Regelwerk, II; 4. und 5. zur Deskriptorenwahl und zur sprachlich-formalen Analyse dieser Deskriptoren zur Wirkung gekommen ist. Die „Begriffszusammenführung“ tritt also in Konkurrenz zur freien Deskriptorenwahl und „Begriffserlegung“ (vgl. B. 14 und Anhang, Abschn. 3).

E.4 Daß in der Norm die Bedeutung und Position der „Begriffszusammenführung“ nicht richtig gesehen werden, geht auch aus dem Hinweis hervor, daß man das Ergebnis, das man mit ihr anstrebt, auch auf andere Weise erzielen könne (II, S. 17; 5.5.): „Wird auf eine geregelte Begriffszusammenführung verzichtet, so muß bei der Suche daran gedacht werden, auch mögliche Zerlegungskomponenten, also übergeordnete Begriffe eines Suchbegriffes mit anzufragen“. Eine von vielen erforderlichen Suchalternativen (Vgl. Anhang, Abschnitt 3.) wäre dann beispielsweise etwa die – sehr ungenaue – Deskriptorenkombination von Beispiel 4 in II, S. 4; 4.2.2. Hierbei ist aber zu beachten, daß

- a) die Einbeziehung derartig merkmalsarmer, allgemeiner Suchalternativen in die Fragestellung zwangsläufig auch entsprechend viel Ballast bringt,
- b) man nach Gutdünken zumindest auf die merkmalsärmsten Suchalternativen verzichten muß, um wenigstens einen Teil dieses Ballastes zu vermeiden. Hierbei ist man aufs Raten und Probieren angewiesen. Dies führt mit hoher Wahrscheinlichkeit zu Verlust, abgesehen davon, daß allein schon das Zusammentragen aller denkbaren Suchalternativen, die man eigentlich berücksichtigen müßte, leicht einen untragbar großen Aufwand verursachen kann.

E. 5 Man muß sich auch vor Augen führen, daß die Anwendung des gesamten Regelwerkes zur Deskriptorenwahl (II, S. 3–9) im ersten Anlauf nur zur Ermittlung von *vorläufigen Deskriptoren*¹⁷ führt. Sie stellen lediglich den Inhalt eines Dokumentes in komprimierter Form dar. Welche Deskriptoren schließlich als *endgültige Deskriptoren* gewählt werden, entscheidet sich erst nach zwei anschließenden, jeweils vielstufigen Prozessen. Es muß nämlich erst noch die formal-sprachliche Lexemanalyse durchgeführt werden (II, S. 10–13; 5.3.–5.4.2.1.2.) und endlich noch die Begriffsanalyse (II, S. 13–16; 5.4.3.–5.4.4.). – Sollte man das Empfinden haben, daß es „zweckmäßig“ (II, S. 17; 5.5.) ist¹⁸, statt der „Begriffserlegung“ die „Begriffszusammenführung“ zu vollziehen, dann ist sogar praktisch das gesamte Regelwerk auf den Seiten 3–16 irrelevant, weil dann verbindlich nach einem vorgegebenen Vokabular indexiert werden muß (vgl. B. 14).

E.6 Die gegenwärtig in der Praxis verbreiteten Indexierungsvarianten wirken frappant einfach und effektiv, wenn man sie mit diesem komplizierten und obendrein noch unzulänglichen Regelwerk der Norm vergleicht.

E.7 Die Norm hebt hervor, daß sie sich nur mit dem Wortschatz einer Indexsprache beschäftigt, nicht jedoch mit der Grammatik, insbesondere der Syntax einer solchen Sprache. Daß die Außerachtlassung jeglicher Grammatik zu erheblichen Schwierigkeiten in der Formulierung eines Regelwerkes führen muß, wenn dieses Anspruch auf Allgemeingültigkeit erhebt, ist verständlich und auch aus der Erörterung der beiden Beispiele deutlich geworden (Vgl. auch Anhang, Abschn. 5). Aber der Normentwurf erweckt in besonders bedenklicher Weise den Eindruck, daß es zu seiner Vervollkommenung *lediglich ergänzender Regeln* bedarf, wenn sich die „Rahmennorm“ als nicht ausreichend erweisen sollte (I, S. 2; 3; II, S. 2; 3.).

E.8 Im Vertrauen auf diese Zusicherung und unter Zurückstellung seiner Bedenken könnte der weniger Erfahrene sich zum Arbeiten nach der Norm entschließen, sofern er, zumindest anfänglich, die niedrige Effizienz dieser Arbeitsweise in Kauf nimmt. Er wird dann viel zu spät bemerken, daß beim Arbeiten nach der Norm die Treffsicherheit des Retrievals immer mehr hinter seinen Erwartungen zurückbleibt und daß dies größtenteils auf die Überlastung seines Wortschatzes mit präkoordinierten Deskriptoren zurückzuführen ist. Diese Überlastung ist dadurch eingetreten, daß der Wortschatz in behelfsmäßiger Weise Aufgaben wahrnehmen mußte, die viel einfacher und wirksamer durch eine Indexsprachen-grammatik erfüllbar gewesen wären. Er wird die Erfahrung machen, daß die syntaktischen Techniken einen viel übersichtlicheren und zuverlässiger handhabbaren Wortschatz ermöglichen und daß die Ausnützung all dieser Möglichkeiten keineswegs eine lediglich ergänzte, sondern eine gänzlich andere Indexierungsweise erfordert als es die Norm empfiehlt. Die Konsequenzen einer späteren Abkehr von der Norm sind schwerwiegend, denn der dann entstehende Informationsspeicher muß ebenfalls mit einer von Grund auf revidierten, und nicht lediglich ergänzten Indexsprache befragt werden. Der dann zurückbleibende Altspeicher muß entweder mit sehr großem Arbeitsaufwand selbst revidiert werden, oder aber er ist nur noch mit einer hinfällig gewordenen, immer mehr der Vergessenheit anheimfallenden Indexsprache befragbar.

F. Schlußbemerkung

F.1 Von den Mängeln des Normenentwurfs konnten im Rahmen dieser Besprechung nur zwei Einzelfälle eingehender erörtert werden. Diese und die weiteren Mängel machen die Empfehlungen der Norm definitiv unannehmbar zumindest im naturwissenschaftlich-technischen Bereich (z.B. Chemie, Pflanzenschutz, Landwirtschaft, Patentwesen). Da aber überhaupt schon die Grundidee verfehlt wirkt, eine für praktisch alle Gebiete und Aufgaben geeignete einheitliche Indexierungsweise in die Form einer Norm zu kleiden, würden sich analoge Limitationen sicherlich auch auf allen anderen Gebieten ergeben, wenn man diese Frage dort entsprechend genau prüfen würde. –

F.2 Aus dieser Sachlage ergibt sich zwangsläufig, daß die Norm keinen Stand der Technik¹⁹ widerspiegelt. Dies gilt ganz besonders für das widersprüchliche, komplizierte und ineffiziente Regelwerk zur Begriffsabgrenzung. Es kann auch nicht Aufgabe einer Norm sein, daß sie den Ideen ihrer Verfasser die Wege ebnet. Soll eine Norm eine Arbeitspraxis vereinheitlichen, so muß sie einen jahrelang praktizierten Stand der Technik widerspiegeln. Allein schon dieses Argument würde ausreichen, um den Normenentwurf abzulehnen.

F.3 Was die Begriffserklärungen im Teil I der Norm anbetrifft, so bedürfen diese ebenfalls nicht der äußeren Gestalt einer Norm. Man entnimmt sie besser der wissenschaftlichen Literatur. Auf einem Gebiet, bei welchem sich Forschung und Entwicklung noch derartig stark im Fluß befinden wie auf informationswissenschaftlichem Gebiet, bedeutet jede Normung eine ernste Behinderung des weiteren Fortschrittes. Normieren ist hier gleichbedeutend mit Sterilisieren.

F.4 Es ist bekannt, daß die Schwierigkeiten, in die man durch die Empfehlungen der Norm geraten wird, erst nach mehrjähriger Arbeit ihre volle Wirkung entwickeln und daß der dann eintretende Schaden kaum noch wieder gutzumachen ist. Er besteht in einer dann rapide abnehmenden Funktionstüchtigkeit der betreffenden Informationssysteme und in einer fortgesetzt ansteigenden Belastung des Informationssuchenden oder seines Beauftragten. Die Recherchenqualität wird demzufolge fortgesetzt absinken.

F.5 Es schien daher notwendig, auf diesem Wege die Öffentlichkeit auf die Sackgassen hinzuweisen, in die sie die Anwendung dieser möglichen Norm DIN 31 623 mit ihren offenkundigen und versteckten Mängeln verleiten würde. Keinesfalls sollte dieser Normentwurf in die internationale Normungsarbeit eingebracht werden, wie bereits beabsichtigt, sondern im Gegenteil: er sollte baldmöglichst zurückgezogen werden.

Anmerkungen:

- 1 Vgl. z.B. die Kritik zur DIN 2330 in (1).
- 2 Siehe auch (2) und R. Fugmann: „Die Grenzen des Geltungsbereiches der Normvorlage DIN 31 623“. Arbeitspapier, Komitee Klassifikations- und Thesaurusforschung (KTF) der Deutschen Gesellschaft für Dokumentation; erhältlich vom Schriftführer des Komitees und vom Verfasser.
- 3 sofern man die Aufgabe des Indexierens darin erblickt, Dokumente für fachliche Fragestellungen leicht und sicher wiederauffindbar zu machen.
- 4 Allerdings beschäftigt sich die Norm in weiten Teilen praktisch ausschließlich mit der komprimierten Darstellung des Inhaltes durch überdies nur vorläufige Deskriptoren (vgl. Abschnitt E.5).
- 5 II, S. 5; 4.2.3.3. Leicht widerlegbar ist auch die dort postulierte direkte Beziehung zwischen Verlässlichkeit und Spezifität, vgl. auch Anhang Abschn. 1.
- 6 Die Gebräuchlichkeit ist obendrein ein viel zu subjektives Kriterium, um für das Indexieren brauchbar zu sein. Wie oft muß ein bestimmter Indexer (oder eine (wie große?) Majorität von Indexern oder auch von anderen Fachleuten?) auf einen Fachausdruck gestoßen sein, bevor sie ihn als „gebräuchlich“ empfinden? (II, S. 5; 4.2.3.5).
- 7 Lochfraß: Eine spezielle Korrosionserscheinung, die lokal scharf begrenzt verläuft und häufig durch Lokalelementbildung oder durch Sauerstoffanreicherung an abgelagerten Fremdkörper-Partikeln ausgelöst wird.
- 8 schließt auch Bruch, Zerreißen, Abrieb, Quetschen usw. ein.

- 9 Dies steht im Widerspruch zu der andernorts gegebenen Empfehlung, die Deskriptoren „in möglichst enger Anlehnung an Titel und Text des Dokumentes zu wählen“ (II, Seite 4; 4.2.3.1.)
- 10 Beispiel: Anstatt „Filter“ „Wasserfilter“ oder vielleicht „Trinkwasserfilter“ oder „Polyamid-Trinkwasserfilter“ oder „Polyamid-Wasserfilter“ oder „Gewebefilter“ usw.
- 11 Dies ist einer der beiden einzigen Hinweise darauf (vgl. I, Seite 3; 5.4.2: „meist DV-Operation“), daß die Norm die Verwendung von Rechenanlagen voraussetzt. Da in dieser Norm ausschließlich Empfehlungen für das Arbeiten mit nichtsyntaktischen Indexsprachen gegeben werden, obwohl doch die Stärke und der Nutzen von Rechenanlagen gerade in den Möglichkeiten zur Anwendung einer Indexsprachensyntax liegen, bedeutet dies zugleich auch die Empfehlung für einen wenig effektiven Einsatz dieser Anlagen.
- 12 Als komplex wird hier jeder Begriff (z.B. Kupferkorrosion) angesehen, der mindestens zwei andere Begriffe (z.B. Kupfer, Korrosion) umfaßt, die ihrerseits im Vokabular vertreten sind. – Die Komplexität eines Begriffes läßt sich für ein Informationsbereitstellungssystem nicht absolut definieren, sondern immer im Bezug auf ein bestimmtes Gebiet und ihr Vokabular. Ein und derselbe Begriff kann für den Fachmann des einen Gebietes komplex sein (z.B. Penicillin für den Chemiker), für den anderen (z.B. für den Landwirt) nicht.
- 13 Zumindest muß man die Erläuterung des Normenentwurfes in II, S. 9; 5.1. wohl in dieser Weise interpretieren. Auf die Mängel der dort gegebenen Definitionen kann hier nicht eingegangen werden.
- 14 und zugleich in einer Form definiert, die viele Zweifelsfälle heraufbeschwört. Ist z.B. in „Alpensteinbock“ „Stein“ oder stattdessen „ein“ ein Lexem? Beide sind eine „sprachliche Einheit“, die als kleinstmöglicher Lexikoneintrag denkbar ist (vgl. II, S. 10; 5.3., vgl. auch die (morphologische) Zerlegung von „Einankerumformner“, in welcher „ein“ als Lexem gilt). Gilt „Blei“ in „Bleistift“ als Lexem, obwohl der Bleistift schon längst nichts mehr mit Blei zu tun hat? Die gegebene Definition ließe sogar zu, „Ei“ als Lexem in „Bleistift“ zu betrachten.
- 15 Der „Lexikonbegriff“ bleibt undefiniert und dürfte auch undefinierbar sein. Er ist demzufolge auch ungeeignet für eine Arbeitsanleitung für das Indexieren. Je nachdem, welches Lexikon man zurate zieht, wird man einen Begriff als „Lexikonbegriff“ einstufen können oder nicht. Wahrscheinlich hat man hier an die Benennungen für Individualbegriffe gedacht, die man der Lexemanalyse entziehen wollte, konnte dies aber nicht ausdrücken, da die Norm den informationswissenschaftlich so wichtigen Unterschied zwischen Individualbegriff und Allgemeinbegriff gar nicht kennt. Die gelegentliche Berufung auf DIN 2330 bedeutet keine Abhilfe, denn dort sind diese Begriffe nur in einer gänzlich untauglichen Weise erklärt (vgl. auch Quelle (1)).
- 16 Wenn man die Analyse von komplexen Begriffen als *unnötige* Erschwerung empfindet, dann sollte man lieber nicht indexieren, sondern stattdessen das Verfassen von Resümees oder Übersichtsartikeln als informationswissenschaftliche Tätigkeit betreiben (vgl. auch B.10).
- 17 Die Verwendung des Wortes Deskriptor für diese Ausdrücke steht auch im Widerspruch zu der für dieses Wort in I, S. 2; 4.1.c) gegebenen Definition.
- 18 genaue Kriterien werden nicht angegeben.
- 19 Daß in der Norm auch altbekannte Tatsachen aufgeführt sind, kann nicht als Nachweis dafür angeführt werden, daß die Norm praktisch angewandt wird, vgl. z.B. Teil II, Seite 2; 4.1.1: „Bei umfangreichen Dokumenten können Titel, Zusammenfassung, Kapitelüberschriften, Bilder . . . für das Verständnis des Dokumenteninhaltes von besonderem Nutzen sein“.

Quellen:

- (1) Fugmann, R.: DIN 2330 oder die genormte Philosophie. In: Nachr. Dok. 28(1977) Nr. 3, (gelbe Seiten, Forum der DGD-Mitglieder)
- (2) Fugmann, R.: Zur Frage der Vereinheitlichung des Indexierens. In: Nachr. Dok. 29(1978) S. 121–126

1. Zur Reproduzierbarkeit (vgl. B2, B3, B9, D7).

Man kann sich leicht vergegenwärtigen, daß die Reproduzierbarkeit des Indexierens keine hinreichende Vorbedingung für das treffsichere Wiederauffinden eingespeicherter Information ist: Eine optimale reproduzierbare Form des Indexierens würde beispielsweise darin bestehen, daß man exakt die von den Autoren verwendeten Wörter als Deskriptoren benutzt, abgesehen von einigen wenigen bedeutungstragenden, vorher festgelegten Wörtern (stop-list). Aber für das Wiederfinden dieser Dokumente mit einer thematischen Fragestellung wäre mit einer lediglich reproduzierbaren Arbeitsweise nicht viel gewonnen:

Kommt in einem solchen Dokument beispielsweise das Wort „Messing“ vor, dann darf man sicher sein, daß ein nochmaliges Indexieren (wenn es wirklich durchgeführt würde) ebenfalls den Deskriptor „Messing“ ergeben würde. Die Forderung der Reproduzierbarkeit ist somit erfüllt.

Ist in einem Dokument der Begriff des Messings durch „Legierung aus Kupfer und Zink“ (in einem bestimmten Mengenverhältnis) ausgedrückt, so würden wiederum hochgradig reproduzierbar die Deskriptoren „Legierung“, „Kupfer“, „Zink“ resultieren.

In einem noch anderen Dokument könnte von „Kupferlegierungen mit Zink“ die Rede sein, ebenfalls mit Angabe des Mengenverhältnisses. Hochgradig reproduzierbar würde dann „Kupferlegierung“, „Zink“ indexiert. Eine noch andere Variante wäre: „Zinklegierung“, „Kupfer“ oder „Tombak“ als spezieller Messingbegriff.

Zum Zeitpunkt der Fragestellung weiß man dann aber nicht, welche und wie viele verschiedene Varianten von Fragen man formulieren muß, um die Dokumente mit dem genannten Legierungsbegriff aufzufinden. Auch ein Thesaurus kann bei einer Indexierungsweise, die sich derart weitgehend nach dem Original richtet, wenig Hilfe bieten. Es ist nämlich unmöglich, in ihm auf die Dauer alle sprachlichen Ausdrucksweisen, unter Einschluß der satz- oder definitionsartigen, zu registrieren und auffindbar zu halten, die im Laufe der Zeit für einen Allgemeinbegriff (oder gar für eine aus Allgemeinbegriffen gebildete Aussage) in den Speicher wandern. Nicht auf die Reproduzierbarkeit der Deskriptorenvergabe kommt es beim Indexieren an, sondern auf deren Rekonstruierbarkeit oder Vorausssehbarkeit.

2. Zum Retrievalmechanismus (vgl. B.7)

Kein Retrievalmechanismus kann und darf eine Fragestellung mit einem *spezifischen* Deskriptor (z.B. Kupferrohrkorrosion, falls im Thesaurus vorhanden) automatisch Antworten allgemeinen Charakters bringen, in denen – *allgemeiner* – nur von Rohrkorrosion oder Metallkorrosion oder Korrosion oder gar nur von Werkstoffverschleiß die Rede ist. Falls allgemeinere Literatur dieser Art gewünscht ist, so kann diese nur durch eine entsprechende Verallgemeinerung der Fragestellung aufgefunden werden. (vgl. auch nachfolgenden Abschnitt).

3. Zur Begriffszusammenführung (vgl. B.5, B.13, E.1, E.2, E.3, E.4)

Wenn die Begriffszusammenführung beim Indexieren unterlassen wird, dann muß man beim Recherchieren bei-

spielsweise die folgenden Suchalternativen berücksichtigen, um die Dokumente von der Art des Beispiels 4 im Teil II, S. 4; 4.2.2. (vgl. B.4, B.13) zu finden:

„Kupfer“, „Rohre“, „Korrosion“	und/oder
„Kupferrohre“, „Korrosion“	und/oder
„Kupferkorrosion“, „Rohre“	und/oder
„Kupferkorrosion“, „Wasserleitungsrohr“	und/oder
„Kupfer“, „Zerstörung“, „Hohlkörper“ usw.	

Wie wenig die – vorläufigen – Deskriptoren (vgl. E.5) für die Begriffszusammenführung geeignet sind, geht besonders deutlich aus der versäumten Vergabe des Deskriptors „Lochfraß“ oder „Lochfraßverhütung“ hervor. Die Fremdkörperpartikeln in diesem Beispiel wurden irrtümlich als unwesentlich angesehen und nicht als vorläufiger Deskriptor ausgedrückt. Die restlichen Deskriptoren des Beispiels geben nun keinerlei Hinweis mehr darauf, daß es sich um diese spezielle Korrosionserscheinung handeln könnte, die typischerweise in horizontal verlaufenden Rohrleitungen an solchen Stellen ausgelöst wird, in denen sich Fremdkörperpartikeln ablagern. Auch die nachfolgend wiedergegebenen Begriffszusammenführungen, falls vom Vokabular verlangt, lassen sich nicht mehr mit den Deskriptoren des Beispiels vollziehen, sondern nur anhand des Originaltextes:

„Kupfer“, „Rohre“	→ „Kupferrohre“
„Polyamid“, „Gewebe“	→ „Polyamidgewebe“
„Wasserleitungsrohr“, „Filter“	→ „Wasserfilter“

4. Begriffskomplexität und Relationenwege (vgl. C.1, C.2, D.7, D.13)

Das Suchen nach den jeweils bestpassenden Deskriptoren in einem Vokabular, wie es beim verbindlichen Indexieren verlangt wird, darf in der Praxis keinen unzumutbaren hohen Aufwand an Zeit und Konzentration erfordern. Soll es gewährleistet sein, daß man beim Suchen mit Sicherheit bis zu den bestpassenden Deskriptoren vorstößt, so darf man sich nicht auf sein Gedächtnis und sein gutes Glück verlassen müssen. Vielmehr müssen alle mehr oder minder gut treffenden (und mehr oder minder nahe verwandten) Deskriptoren für ein Thema durch vorgebahnte Wege miteinander verbunden sein, die man beim Suchen nur zu verfolgen braucht, um schließlich bis zum bestpassenden Deskriptor hingeleitet zu werden. In einem Wortschatz, der für das verbindliche Indexieren bestimmt ist, sind diese „Relationenwege“ durch die hierarchische Anordnung der Deskriptoren vorgebahnt, sowie auch durch die ins Vokabular eingetragenen Verweisungen. Sind diese Relationswege sehr lang und muß man beim Verfolgen dieser Wege vielen Verzweigungen nachgehen (wobei sich die Abzweigung vielleicht fortgesetzt vielfach weiterverzweigen), dann wird man in der Praxis oft das Ziel der Suche, nämlich die bestpassenden Deskriptoren, nicht erreichen: Nichtverbindliches Indexieren ist die zwangsläufige Folge.

Will man verhindern, daß sich im Vokabular ein solcher Zustand entwickelt, dann darf man nicht jeden beliebigen Deskriptor für das Vokabular zulassen. Insbesondere müssen diejenigen Deskriptoren aus dem Wortschatz möglichst ferngehalten werden, die zu einer intensiven Vernetzung und Verzweigung der Relationenwege führen würden (siehe oben, Abschn. 5).

Man kann sich leicht vergegenwärtigen, daß die Vernetzung eines Wortschatzes vor allem durch die präkoordinierten Deskriptoren für komplexe¹² Begriffe zustande kommt.

Wie groß der Beitrag eines Deskriptors zur Vernetzung der Relationenwege ist, hängt ausschließlich davon ab, wie komplex er ist, d.h. wie viele andere Begriffe er einschließt, die ebenfalls im Vokabular durch Deskriptoren vertreten sind. Insbesondere ist es für die Wortschatzzeichnung eines Begriffes gleichgültig, wie ein solcher Begriff sprachlich ausgedrückt oder ausdrückbar ist. Betrachtet man beispielsweise den Begriff der

„Korrosion von Rohren aus Kupferlegierungen mit einem Gehalt von 40% Zink“,
so ändert sich nichts am komplexen Charakter dieses Begriffes und an seiner geringen Eignung für den Wortschatz, wenn man ihn durch

„Messingrohrkorrosion“
ausdrückt. Begriffe, die im Rahmen eines bestimmten Vokabulars ausgesprochen komplex wirken, und die man daher dort im Interesse der verbindlichen Anwendbarkeit dieses Vokabulars vermeiden sollte, können sprachlich sehr einfach ausdrückbar sein:

Rohrzucker, Morphem, Aquaplaning, Messingrohre, Silikose, Rost usw.
Umgekehrt gibt es für viele geläufige Fachbegriffe (noch) keine Wörter:

„Der Korrosion unterliegendes Material“.
„Unfallgefahr durch Ölsuren auf der Fahrbahn“.
„Geschwindigkeit des Zerfalls von Pflanzenschutzmitteln unter der Einwirkung von Licht und Feuchtigkeit“.

Es ist demzufolge eine fruchtlose und abwegige Übung, wenn man die Wortschatzzeichnung von Begriffen und die evtl. Notwendigkeit zu ihrer Zerlegung von der Zahl der Wortstämme oder Lexeme abhängig macht, durch die der Begriff ausgedrückt wurde (oder ausdrückbar ist).

5. Zur Syntax (vgl. C.2, E.7, E.8)

Verschiedene unrichtige Feststellungen und Regeln des Normentwurfs beruhen auf dem Irrtum, daß die zusammenhanglose Aufzählung von begrifflichen Komponenten eine adäquate Wiedergabe vom Inhalt eines Dokumentes sein könne. Ein Beispiel hierfür ist II, S. 13; 5.4.3.1.: „Ziel der Begriffsanalyse ist es, die Einzelbegriffe zu ermitteln, die in ihrer Kombination (Postkoordination) genau den Begriffsinhalt des ursprünglichen Begriffes wiedergeben“. Sind aber diese „Einzelbegriffe“ lediglich zusammenhanglos aufgezählt, dann können sie nur dann den ursprünglichen Begriff *genau* wiedergeben, wenn außer ihm im gleichen Dokument keine weiteren Begriffe indexiert worden sind. Andernfalls lassen sich aus einer solchen zusammenhanglosen Aufzählung von Einzelbegriffen (z.B. aus den Komponenten verschiedener Metall-Legierungen) viele verschiedene falsche Kombinationen (z.B. Legierungsbegriffe) herauslesen. Es wäre die Aufgabe einer Indexsprachgrammatik, insbesondere der Syntax einer solchen Grammatik, diese Zusammenhänge darzustellen. Ist eine solche Syntax verfügbar, dann lassen sich viele präkoordinierte Deskriptoren im Wortschatz vermeiden, ohne daß man an Wiedergabetreue verliert.

Classification Teaching in India

Syllabus for advanced courses in classification and indexing in India (Courses at Masters and Post-Master's level in Information Science in India; Courses conducted at DRTC, Bangalore, INSDOC, New Delhi, Delhi University, Madras University, Bangalore University).

1. Study of subjects: Universe of subjects, Modes of formation of subjects, Recognition of interdisciplinary subjects, Transdisciplinary subjects and their models.

2. Mapping of subjects in general classification schemes. Philosophical models and their derivations to classification. The principles in mapping these subjects. Facet analysis. The objectives of facet analysis. Postulates and principles of facet analysis. Theories of faceted classification. Fundamental categories and operators.

3. Concept of array, chain, and collateral structures. Presentation of ideas in paradigmatic structures. Principles for helpful sequence.

4. Representation of hierarchical and non-hierarchical structures/relations. Typology of relations among ideas. The role of this typology in semi-syntactic representation of ideas and subjects in expressions for indexing.

5. Faceted classification schemes: UDC and CC. The environmental restrictions of these classification schemes. History, basic philosophy of these schemes in three planes: Idea plane, Verbal plane and Notational plane. Study of typology of notations. Hierarchical and Non-hierarchical notations. Practical assignments in assigning UDC and CC numbers.

6. Theoretical basis of indexing and indexing language. Evaluation of subject indexing since Cutter, Kaiser, Ranganathan. Chain indexing, POPSI, and PRECIS. Limitation of these indexing languages. Practical in indexing; by chain indexing, POPSI, and PRECIS.

7. Study of Thesaurifacet. Role and structure of thesaurus in information retrieval. Steps in designing of classification schemes. Steps in designing thesaurus. Use of thesaurus in indexing. Post-co-ordinate indexing and pre-conditioned indexing including the early attempts.

8. Searching information. Search language, Search mechanisms and strategies, Use of roles and links, Boolean logic, Use of class numbers, PRECIS headings, and POPSI headings in retrieval of information.

9. Computer based information systems. Classification, indexes, and thesaurus for the same. Vocabulary control devices. Generation of these by using computers.

10. Evaluation of indexing and searching. Empirical approach and theoretical approach. Recall and precision studies. Qualitative and quantitative evaluations. Cost-effectiveness and cost benefit analysis.

11. Interdisciplinary approach to classification and indexing science. Principles of logic and its application to classification process. Linguistic theories to classification, Psychological theories and classification. Behavioural aspects of users vis-a-vis classification.

12. Classification theory as a scientific discipline. Its evaluation and status. Glossaries in the field of classification and indexing.

13. National and international approaches to classification. FID/CA classification committee of FID (1950–1962). Dorking conference (First International Study Conference in Classification Research, (1957). FID/CR Classification Research Committee (1963–72) (Denmark). Elsinore Conference (Second International Study Conference in Classification Research) (Elsinore) (1964) FID/CR (1973–) (India). Bombay Conference (Third International Study Conference on Classification Research) (1975).

14. Periodical publications supporting classification and indexing research.

Information processing and management (USA), Journal of American Society for Information Science (USA), Journal of Documentation (UK), International Classification (Germany), *Annals of Library Science and documentation* (India), *Library Science with a slant to documentation* (India).

15. Areas for research in classification and indexing. New directions in classification. Project in the application of a model of classification and indexing in information systems.

(Ed. note: We gratefully acknowledge this communication through the kindness of Prof. Gopinath, Documentation Research and Training Centre (DRCT), Bangalore.)

Informatics 5 – “Analysis of Meaning”

The 5th Conference of the Aslib Co-ordinate Indexing Group (now Informatics Group) was held in Oxford on March 21–23, 1979. It was co-sponsored by the British Computer Society. There were 27 papers presented – too many and not enough discussion time left. The first keynote address was by *Aaron Sloman*: “Semantics: a viewpoint from Artificial Intelligence”, (AI). He dealt with AI in relation to language understanding, picture processing and vision, interpretation of music and theorem proving. (His book, *The Computer Revolution in Philosophy*, Harvester Press, 1978, is of interest.) The second key note paper by *A. J. Cain*: “Taxonomic methods in biology”, reviewed the problem of classifying by genus, species and even by whole populations. The short paper by *P. J. Best* (“Data interpretation using an elliptical display”) could be of possible use to people dealing with quantitative business analysis. The paper by *F. Parker-Rhodes* (“Information processing – restricted or elaborated?”) is basically a criticism of Bernstein’s restricted and elaborate codes and refers to his recent book on Inferential Semantics. The paper by *C. R. S. Marsden* (“False tautologies, unthinkable worlds and their impossible mental objects”) is a study of Leibniz and his theory of Monads, however, purely of theoretical importance. – Instead of G. N. Leech, *F. E. Knowles* gave an extended paper on “Recent Soviet work on computer techniques for representing natural language meaning”. He studied the work on Machine Translation in the Soviet Union since the 1950s concerning also the work of Mel’cuk, there called ‘engineering linguistics’. The English-Russian dictionary MARS is programmed in COBOL; it includes thesaurus information and an algorithm of 99% lemmatization of

Russian text. The paper by *R. L. Johnson* and *H. L. Somers* is called: “PTOSYS”. (Ptosys means ‘case’. The authors claim it to be an interactive system for understanding texts of esoteric interest, particularly for mapping surface cases onto deep cases. *P. Galloway* talked on “Yngve’s depth hypothesis and the structure of narrative detective fiction”. The *P. W. Williams* (“The use of computer language principles to clarify the meaning of written English”) paper tried to transform English written language into a form resembling computer language in order to cut down imprecision, redundancy, gestures, tonality, rhythm and pauses. *A. M. Spector* (“Some problems in the analysis of multi-segment texts”) studied the coherence of texts. His work is based on Frege. He presented language as a problem solving process. *M. Januszewicz*’s paper on “The analysis of meaning” gave an example of the use of the UDC in an agricultural library. The keynote paper by *Y. Wilks* (“Semantics in Artificial Intelligence”) was basically a critique of the Sussex School, Department of Experimental Psychology and the USA School led by Woods of Harvard. *E. Hollnagel* (of Denmark) (“The relation between intention, meaning and action”) gave a very good cognitive paradigm of communications. *B. Kostrewski*’s paper (“Structural considerations for the derivation of application linked reference languages for medical information systems”) mentioned also the Standard Nomenclature of Medicine. *C. Bell* and *K. P. Jones* dealt with practical indexing in “Back-of-the-book indexing: a case for the application of Artificial Intelligence”; he was particularly concerned with ‘aboutness’. *G. P. Zarri* (of CNRS, Paris) (“Artificial Intelligence and information retrieval: a look at the RESEDA Project”) gave a very intelligent paper on advanced question answering systems dealing with metalanguage and inference procedures. *V. Verdier* (“The PRECIS-Translingual Project”) presented the PRECIS methodology as extended to other languages than English. *W. Goshawke* “Number language and information retrieval”) found little echo; the speaker apparently thought all words could be reduced to one number. He was reminded that this had already been done for Chinese telegraph number codes. *K. Sparck Jones* (“Problems in the representation of meaning in information retrieval”) concluded that AI was too refined for document retrieval. *S. E. Robertson* (“Between aboutness and meaning”) followed that AI and linguistics being inappropriate at the moment for document retrieval; this should find its own objectives. *G. Lubbock* of Rome (“The achievement of meaning in IR: the problem and some proposals”) presented a very practical problem of terminology for economic activity throughout Western Europe (SISDATA). *J.-C. Gardin* (“On the relation of question-answering systems and various theoretical approaches to the analysis of text”) showed three aspects of convergence: the increasing role of semantics in linguistic analysis since Fodor 1963, the logical versus semantical organisation in Fillmore’s Case Grammar and the convergence between linguistics and information retrieval. *P. L. Noerr* (“Information navigation”) presented a practical paper on document retrieval systems. *H. Brookes*, *N. J. Belkin* and *R. N. Oddy* (“Representing and classifying anomalous states of knowledge”) presented an untypical tape recorded interview with on-line system users, valuable for strategy improvement. *I. M.*

Pigott of EEC Luxembourg ("Theoretical options and practical limitations of using semantics to solve problems of natural language analysis and machine translation") related experiences with the Systran machine translation system. *K. V. Hanford, D. J. H. Brown, W. J. Black* ("EDOR: an interactive system for easy document retrieval") presented their robust, flexible and easy-to-use system linking Teesside Polytechnic with the British Deal Information Service. *J. M. Buchanan* ("Perspectives in a polluted world — tools and procedures for information management") showed that a thesaurus can be used also by researchers to index their own material.

(According to notes from) A. J. Dickson

Matter and Methods of Subject Access to Knowledge

Subsequent to a First Seminar on "Classification and Education" (see I.C. 6 (1979) No. 1, p. 37–38) a Second Seminar was held at the Bibliothekar-Lehrinstitut, Köln, 28–29 March 1979 on the topic "Lehrinhalte und Lehrmethoden des Faches Sacherschließung" (attempted translation in the heading). The seven papers presented were mostly results from questionnaire evaluations aiming at a survey on the matter, objectives and methods of classification teaching at the different library schools in Germany (Federal Republic) for differing levels of activities (e.g. for public libraries or scientific libraries, for documentation and information science). The first session was devoted to "Lehrinhalte und Lehrziele im Fach Sacherschließung". *B. Lewe* presented the situation "aus der Sicht der Öffentlichen Bibliotheken", *Franz Fischer* "aus der Sicht der wissenschaftlichen Bibliotheken". *B. Kirchhoff* gave a survey on "Lehrinhalte des Faches Dokumentation im Hinblick auf die bibliothekarische Sacherschließung". The second session concerned itself with teaching objectives and teaching aids. *G. Greiner* gave a survey on the syllabus of the German Documentation Society's Documentation Training Institute regarding the knowledge in classification and indexing which the medium level of documentalists (diplomierte Dokumentare) are supposed to have; *G. Heinrich* had collected information on the teaching aids (especially textbooks) used in this field altogether; *R. Klingbeil* outlined the teaching objectives from the part of the library practice in scientific libraries and *Ch. Erpf* did a similar outline from the part the public libraries. — All of the papers demonstrated the great variety in theory and practice at all of these schools. The lack of suitable teaching aids was very much felt. Rather spontaneously it was decided to form a group concerned with establishing a collection of samples from existing classification schemes and thesauri which could be regarded as a common basis for teaching in this field. The group will convene in October 1979. The papers of the second symposium will be published together with some of the first one by the Bibliothekar-Lehrinstitut, Köln.

I. Dahlberg

Pattern Recognition (PR) and Artificial Intelligence (AI)

A total of over 100 specialists in PR and AI were brought together in a Workshop on Pattern Recognition and Artificial Intelligence of the IEEE Computer Society, Technical Committee on Machine Intelligence and Pattern

Analysis, April 12–14, 1978, Princeton, N. J. Some 37 papers were submitted, a selection of them concerning analysis of images, analysis of speech and certain general algorithms is included in the second issue of the new journal "IEEE Transactions on Pattern Analysis and Machine Intelligence" 1 (1979— (p. 125–235). In the Preface to this issue, Y. T. Chien and Th. Pavlides from Princeton University give a short account of the contents of these selected papers. There are also a number of articles on cluster-analysis. The authors: "a subject which is of equal importance to both PR and AI. . .". The question of whether there exist substantial common methodologies between AI and PR was the subject of discussion and variance in opinion at the Workshop. The lack of maturity of these fields may be a reason for an impossibility to define their scopes, as this may be done, e.g. for physics and chemistry. Researchers should be aware of this problem.

1979 Annual General Meeting of The Classification Society, European Branch

At this years' Annual General Meeting on April 30, 1979 at 2 pm in the Marcus Beck Library, Royal Society of Medicine, London, a lecture was held by Dr. Harvey A. Skinner, Addictions Research Foundation, Toronto, Canada, on: "Classification of abnormal behaviour: critical issues, a model and some empirical evidence". He showed developments in psychiatric classification and explained a dimensional model of psychopathology with reference to Sneath's vector model of diseases.

French Seminar on "Ordre et Classification"

At the Centre de Mathématique Sociale (54, Blvd. Raspail, 75006 Paris) a seminar on "Order and Classification" took place which was held on April 23, May 7 and May 21, 1979. The following lectures were given: F. de la Vega: Une généralisation probabiliste de l'opération de fermeture transitive. — C. Dupont-Gateland: Une méthode de classification non-hiérarchique sur des données qualitatives et quantitatives. — J. F. Marcotorchino, P. Michaud: Optimisation en analyse des données ordinales; méthodes heuristiques.

Chemosystematics: Principles and Practice

This was the topic of an international symposium of the Systematics Association held at the University of Southampton, July 10–12, 1979. In its invitation it was stated that "the 12 years since the Association's pioneer symposium "Chemotaxonomy and Serotaxonomy" have seen a massive expansion of chemosystematic work". 21 papers and one poster session foreseen demonstrate some of the development. The papers are to be published through Academic Press. The conference was concerned with four special topics: (1) Secondary substances, (2) Proteins and other large molecules, (3) Analysis of chemical data, and (4) Principles and practice. For further information please contact Dr. F. A. Bisby (Chemosystematics Symposium), Biology Department, Bldg. 44, University, Southampton SO9 5NH, England.

Dictionary Society of North America, 1979 Meeting

The Biennial Meeting took place at London, Ontario, Canada, June 28–30, 1979 with the following 13 papers on the program: *W. S. Avis*: Trade names, lexicographers, and the law. — *St. A. Kingsbury*: Semantic categories and regional distribution of key words in American proverbs and proverbial sayings of the American Dialect Society. — *R. H. Secrist*: Is respelling really necessary? — *K. Roe*: Progress report on a dictionary of eponymic and descriptive names of theoretical concepts in biology. — *A. W. Read*: Lexicographers are human: an account of personal bias in definitions. (Keynote address). — *W. J. Cameron*: Browsing files for dictionary scholars. — *J. Hirshberg*: Spoken usage, status labels, and DARE. — *D. Sherman*: A computer archive of machine readable dictionaries. — *D. L. Gold*: The ordering of lexemes in a dictionary. — *P. Drysdale*: Dictionary etymologies: what? why? and for whom? — *D.-H. Nguyen*: Bilingual lexicography in Vietnam: the state of the art. — *A. C. Amos*: Lexical treatment of the function words in the dictionary of Old English. — *D. S. Marshall*: Lexica for Polynesia: problems, values and lessons learned in bilingual dictionary making. — *Sh. Kuhn*: The Middle English Dictionary: four centuries of English life and language. For further information please contact Dr. William J. Cameron, The University of Western Ontario, London, Ont., Canada.

Classauri in Information Work

The Documentation Research and Training Centre (DRTC), (Indian Statistical Institute), Bangalore has taken up a project for designing an indexing vocabulary which would incorporate all the essential features of a scheme of faceted classification together with those of a systematic thesaurus. The classified/systematic part will be supplemented by an alphabetical index in which each term would occur only once with its address. This new type of compilation has been called "Classaurus" (= Faceted Classification Scheme + Thesaurus). Experiments on this have been conducted in DRTC since 1967. A "Classaurus" on Library and Information Science was designed in 1968. Since then, it has been taught regularly in the course conducted by DRTC. This year (1979) a large scale project has been taken up for designing "Classauri" on various subjects. A "Classaurus" ensures considerable economy in presentation, and ease in understanding and application. It can be conveniently used for pre-coordinate indexing, post-coordinate indexing, and for arranging documents on shelves, and surrogates in files. For further information, please contact G Bhattacharyya, Professor, Documentation Research and Training Centre, Indian Statistical Institute, 31 Church Street, Bangalore 560 001, India.

From: FID/CR Newsletter 6 (1978) No. 4, p. 35

Conference on Pattern Recognition

The British Pattern Recognition Association will hold a conference in Oxford, 9–11 January 1980 and invites to hand in papers on the following topics:

Pattern recognition theory (statistical and syntactic methods, clustering)
Processing and recognition of speech and other wave-forms
Image and line drawing processing
Applications of pattern recognition (biomedical, industrial, social and physical sciences, remote sensing, etc.)
Scene analysis
Robotics
Artificial intelligence
Hardware and software systems.

For contributed papers three copies of an abstract of 300–500 words should be sent to Dr. Josef Kittler, Nuclear Physics Laboratory, Keble Road, Oxford OXI 3RH, England, not later than 1 July 1979.

A New Order for the Names of Bacteria

At the 12th World Congress of Microbiology in Munich the International Committee for Systematic Bacteriology (ICSB) convened for an important decision. It had worked for some time toward a new nomenclature for bacteria. This will become valid by Jan. 1, 1980. From then on descriptions of newly detected bacteria will be published only in "International Journal of Systematic Bacteriology".

Numerical Data Banks: 3,000,000 Statistical Items now available for Italy

Slamark International, an Italian economic and marketing research consultant firm, has recently developed a statistical information system. SISDATA, accessible on-line through the ADP Time-sharing Network.

The numerical file, which covers all major areas of the Italian economy, amounts currently to a total of 3 million figures and extends back to 1971.

Raw data are supplied in tabular form with complete bilingual descriptions (English/Italian). A specific command language, designed by Slamark, permits handling the statistics in order to obtain the aggregation of parameters, the suppression of rows/columns, the switching of row/columns, etc. Moreover, statistical programmes from the ADP Library can be applied to SISDATA for forecasting, regression and correlation analysis, etc.

SISDATA is constantly up-dated, drawing on official and semi-official sources, at a rate of some 600,000 figures a year.

A technical and scientific committee composed of Italian academicians and experts has collaborated in the elaboration of the theoretical approach to SISDATA, since its conception in 1976 and has guided its practical application throughout the various stages of the creation of the data bank.

SLAMARK International will be happy to supply further information and descriptive material on SISDATA. Inquires should be addressed to: SISDATA, SLAMARK International, Via Ignazio Guidi, 4, 00147 ROME. Tel. (06) 5140171/476 (Telex n° 613408)