

Using the values derived in Figure V,3 the following probabilities can be derived.

$$(1) P(E_r/V_r) = P(E_r/C_r) P(C_r/V_r) + P(E_r/C_{\bar{r}}) P(C_{\bar{r}}/V_r)$$

$$(2) P(E_r/V_{\bar{r}}) = P(E_r/C_r) P(C_r/V_{\bar{r}}) + P(E_r/C_{\bar{r}}) P(C_{\bar{r}}/V_{\bar{r}})$$

$$(3) P(R_r/V_r) = P(R_r/E_r) P(E_r/V_r) + P(R_r/E_{\bar{r}}) P(E_{\bar{r}}/V_r)$$

$$(4) P(R_r/V_{\bar{r}}) = P(R_r/E_r) P(E_r/V_{\bar{r}}) + P(R_r/E_{\bar{r}}) P(E_{\bar{r}}/V_{\bar{r}})$$

$$(5) P(S_r, R_r/V_r) = P(S_r/V_r) P(R_r/V_r)$$

$$(6) P(S_r, R_r/V_{\bar{r}}) = P(S_r/V_{\bar{r}}) P(R_r/V_{\bar{r}})$$

$$(7) P(R_r/V_r) + P(S_{\bar{r}}, R_r/V_r) + P(S_r, R_r/V_r) = 1$$

$$(8) P(R_{\bar{r}}/V_{\bar{r}}) + P(S_{\bar{r}}, R_r/V_{\bar{r}}) + P(S_r, R_r/V_{\bar{r}}) = 1$$

The notation $P(S_r, R_r/V_r)$ indicates the probability that the system has classified the document as relevant and that the screener has also classified it as such, given that the document is, in fact, relevant with respect to the verbalized request.

The model has the following important features:

- a) It provides a measure of performance of the searching subsystem based on the performance of the components.
- b) The effect of a change in performance of a component on the performance of the subsystem can be evaluated.
- c) The customary measures of recall and precision can be defined in terms of the probabilities thus:

$$\text{Recall Ratio: } P(S_r, R_r/V_r)$$

$$\text{Precision Ratio: } \frac{N_r \cdot P(S_r, R_r/V_r)}{N_r \cdot P(S_r, R_r/V_r) + N_{\bar{r}} \cdot P(S_r, R_r/V_{\bar{r}})}$$

$$N_r \cdot P(S_r, R_r/V_r) + N_{\bar{r}} \cdot P(S_r, R_r/V_{\bar{r}})$$

Where N_r is the number of documents in the file relevant to the verbalized request and $N_{\bar{r}}$ the number nonrelevant.

Using the values derived in Figure V,3 the following probabilities can be derived.

$$(1) P(E_r/V_r) = P(E_r/C_r) P(C_r/V_r) + P(E_r/C_{-r}) P(C_{-r}/V_r)$$

$$(2) P(E_r/V_{-r}) = P(E_r/C_r) P(C_r/V_{-r}) + P(E_r/C_{-r}) P(C_{-r}/V_{-r})$$

$$(3) P(R_r/V_r) = P(R_r/E_r) P(E_r/V_r) + P(R_r/E_{-r}) P(E_{-r}/V_r)$$

$$(4) P(R_r/V_{-r}) = P(R_r/E_r) P(E_r/V_{-r}) + P(R_r/E_{-r}) P(E_{-r}/V_{-r})$$

$$(5) P(S_r, R_r/V_r) = P(S_r/V_r) P(R_r/V_r)$$

$$(6) P(S_r, R_r/V_{-r}) = P(S_r/V_{-r}) P(R_r/V_{-r})$$

$$(7) P(R_r/V_r) + P(S_r, R_r/V_r) + P(S_{-r}, R_r/V_r) = 1$$

$$(8) P(R_{-r}/V_{-r}) + P(S_r, R_{-r}/V_{-r}) + P(S_{-r}, R_{-r}/V_{-r}) = 1$$

The notation $P(S_r, R_r/V_r)$ indicates the probability that the system has classified the document as relevant and that the screener has also classified it as such, given that the document is, in fact, relevant with respect to the verbalized request.

The model has the following important features:

- a) It provides a measure of performance of the searching subsystem based on the performance of the components.
- b) The effect of a change in performance of a component on the performance of the subsystem can be evaluated.
- c) The customary measures of recall and precision can be defined in terms of the probabilities thus:

$$\text{Recall Ratio: } P(S_r, R_r/V_r)$$

$$\text{Precision Ratio: } \frac{N_r \cdot P(S_r, R_r/V_r)}{N_r \cdot P(S_r, R_r/V_r) + N_{-r} \cdot P(S_r, R_r/V_{-r})}$$

$$N_r \cdot P(S_r, R_r/V_r) + N_{-r} \cdot P(S_r, R_r/V_{-r})$$

Where N_r is the number of documents in the file relevant to the verbalized request and N_{-r} the number nonrelevant.

d) The model measures effectiveness rather than efficiency, since no costs are introduced.

The conditional probabilities listed above can be summarised in 2 x 2 tables as shown in Figure V.4 below.

<u>Cumulative through encoding stage</u>		
Relevance with respect to verbalized request	E_r^-	E_r
V_r	$P(E_r^-/V_r)$	$P(E_r/V_r)$
V_r^-	$P(E_r^-/V_r^-)$	$P(E_r/V_r^-)$
<u>Cumulative through system output</u>		
Relevance with respect to verbalized request	R_r^-	R_r
V_r	$P(R_r^-/V_r)$	$P(R_r/V_r)$
V_r^-	$P(R_r^-/V_r^-)$	$P(R_r/V_r^-)$
<u>Cumulative through screening</u>		
Relevance with respect to verbalized request	S_r^-	S_r
V_r	$P(S_r^-, R_r/V_r)$	$P(S_r, R_r/V_r)$
V_r^-	$P(S_r^-, R_r/V_r^-)$	$P(S_r, R_r/V_r^-)$

Figure V.4 Cumulative Probabilities Through Various Stages of the
Retrospective Search

5.3 DATA COLLECTION

In order to implement an evaluation of the retrospective searching system of Ciba-Geigy's library in terms of the theory outlined above, it was necessary to collect relevant data in a systematic manner. Since the model is statistically based, it was considered that at least 100 sample observations would be required for a meaningful application of the model. A standard form was devised which permitted the responses at each stage of an inquiry to be noted by merely placing a tick against the relevant response. An example of the standard form is shown in Figure V.5.

Query	Date
	Time
.....	
.....	
<u>Verbalized Request</u>	
Relevant	Yes No
<u>Coder's Interpretation</u>	
Relevant	Yes No
<u>Indexed Request</u>	
Relevant	Yes No
<u>System's Response</u>	
Relevant	Yes No
<u>Screening's Judgement</u>	
Relevant	Yes No

Figure V.5 The Standard Form Utilized for Data Collection

The completion of these forms was not always found to be a straightforward procedure as the decision as to relevance or nonrelevance on a particular point required judgement and was, therefore, to an extent, subjective. This difficulty, however, was substantially diminished by the fact that all the queries were handled by a staff of only two, thereby enabling a measure of consistency to be achieved.

The final number of inquiries which were followed through to finality was 121 and these results were used for the statistical analysis in terms of the model.

5.4 STATISTICAL ANALYSIS

The estimates of conditional probabilities were reduced to tabular form as illustrated above and yielded the values shown in Figure V,6 below. The tilde ($\sim$) over the noted probabilities indicates that they are probabilities estimated from a sample.

A. Errors in interpreting verbalized request

Verbalized request	<u>Coder's interpretation</u>	
	C_r^-	C_r
V_r	(1) $\tilde{P}(C_r^-/V_r)$ = 0,03	(2) $\tilde{P}(C_r/V_r)$ = 0,97
V_r^-	(3) $\tilde{P}(C_r^-/V_r^-)$ = 0,67	(4) $\tilde{P}(C_r/V_r^-)$ = 0,33

Figure V,6

Cont/...

B. Errors in encoding the request (deficiencies in language)

		<u>Encoded request</u>	
Coder's interpretation		E_r^-	E_r
	C_r	(5) $\tilde{P}(E_r^-/C_r)$ = 0	(6) $\tilde{P}(E_r/C_r)$ = 1,0
	C_r^-	(7) $\tilde{P}(E_r^-/C_r^-)$ = 0,62	(8) $\tilde{P}(E_r/C_r^-)$ = 0,38

C. Errors due to indexing

		<u>Retrieved by system</u>	
Encoded request		R_r^-	R_r
	E_r	(9) $P(R_r^-/E_r)$ = 0,05	(10) $P(R_r/E_r)$ = 0,95
	E_r^-	(11) $P(R_r^-/E_r^-)$ = 0,60	(12) $P(R_r/E_r^-)$ = 0,40

Figure V,6

Cont/...

D. Errors due to screening

Verbalized request	Screened output	
	S_r^-	S_r
V_r	(13) $P(S_r^-/V_r)$ = 0,03	(14) $P(S_r/V_r)$ = 0,97
V_r^-	(15) $P(S_r^-/V_r^-)$ = 0,64	(16) $P(S_r/V_r^-)$ = 0,36

E. Probability of encoding relevance for documents relevant and non-relevant with respect to the verbalized request

(1)	0,03	X	(7)	0,62	=	(17)	0,02	
(2)	0,97	X	(5)	0	=	(18)	0	
			(17) + (18)		=	(19)	0,02	$P(E_r^-/V_r)$
			1,0 - (19) 0,02		=	(20)	0,98	$P(E_r/V_r)$
(3)	0,67	X	(7)	0,62	=	(21)	0,42	
(4)	0,33	X	(5)	0	=	(22)	0	
			(21) + (22)		=	(23)	0,42	$P(E_r^-/V_r^-)$
			1,0 - (23) 0,42		=	(24)	0,58	$P(E_r/V_r^-)$

F. Probability of system response for documents relevant and non-relevant with respect to the verbalized request.

(19)	0,02	X	(11)	0,06	=	(25)	= 0,12
(20)	0,98	X	(9)	0,05	=	(26)	= 0,05
			(25) 0,12 + (26) 0,05		=	(27)	= 0,17 $P(R_r/V_r)$

Figure V,6

Cont/...

D. Errors due to screening

Verbalized request	Screened output	
	S_r^-	S_r
V_r	(13) $P(S_r^-/V_r)$ = 0,03	(14) $P(S_r/V_r)$ = 0,97
V_r^-	(15) $P(S_r^-/V_r^-)$ = 0,64	(16) $P(S_r/V_r^-)$ = 0,36

E. Probability of encoding relevance for documents relevant and non-relevant with respect to the verbalized request

(1)	0,03	X	(7)	0,62	=	(17)	0,02	
(2)	0,97	X	(5)	0	=	(18)	0	
			(17) + (18)		=	(19)	0,02	$P(E_r^-/V_r)$
			1,0 - (19) 0,02		=	(20)	0,98	$P(E_r/V_r)$
(3)	0,67	X	(7)	0,62	=	(21)	0,42	
(4)	0,33	X	(5)	0	=	(22)	0	
			(21) + (22)		=	(23)	0,42	$P(E_r^-/V_r^-)$
			1,0 - (23) 0,42		=	(24)	0,58	$P(E_r/V_r^-)$

F. Probability of system response for documents relevant and non-relevant with respect to the verbalized request.

(19)	0,02	X	(11)	0,06	=	(25)	= 0,12
(20)	0,98	X	(9)	0,05	=	(26)	= 0,05
			(25) 0,12 + (26) 0,05		=	(27)	= 0,17 $P(R_r/V_r)$

Figure V,6

Cont/...

$$\begin{array}{rclclcl}
 1.00 & - & (27) \ 0,17 & = & (28) \ 0,83 & P(R_+/V_+) \\
 (23) \ 0,42 & \times & (11) \ 0,60 & = & (29) \ 0,25 & \\
 (24) \ 0,58 & \times & (9) \ 0,05 & = & (30) \ 0,03 & \\
 (29) \ 0,25 & + & (30) \ 0,03 & = & (31) \ 0,28 & P(R_+/V_+) \\
 1.00 & - & (31) \ 0,28 & = & (32) \ 0,72 & P(R_+/V_+)
 \end{array}$$

G. Probability of screened selection for documents relevant and non-relevant with respect to the verbalized request

$$\begin{array}{rclclcl}
 (14) \ 0,97 & \times & (28) \ 0,83 & = & (33) \ 0,81 & P(S_+, R_+/V_+) \\
 (28) \ 0,83 & - & (33) \ 0,81 & = & (34) \ 0,02 & P(S_+, R_+/V_+) \\
 (16) \ 0,36 & \times & (32) \ 0,72 & = & (35) \ 0,26 & \\
 (32) \ 0,72 & - & (35) \ 0,26 & = & (36) \ 0,46 &
 \end{array}$$

H. Summary

Verbalized request	Not retrieved		Retrieved	
Relevant documents				
Prior to screening	(27)	0,17		
After screening	(34)	0,02	(33)	0,81
Non-relevant documents				
Prior to screening	(31)	0,28		
After screening	(36)	0,46	(35)	0,26

I. Measures of recall and precision

$$\text{Recall} \quad \text{Item (33)} \quad = \quad 0,81$$

Precision

$$\text{Number in file identified as relevant} \quad (N_+) \quad = \quad 539 \text{ (est)}$$

$$\text{Total non-relevant documents in file} \quad (N_-) \quad = \quad 6500 \text{ (est)}$$

Figure V,6

Total retrieved	$N_r \times (33) 0,81$	=	437
Plus	$N_r \times (35) 0,26$	=	1690
	Equals		2127

$$\text{Precision ratio} = \frac{N_r \times (\text{Item } 33)}{\text{Total retrieved}} = \frac{539 \times 0,81}{2127} = \underline{\underline{0,21}}$$

Figure V,6 Worksheet for Calculating Output from the Retrospective Searching Model

5.5 OBSERVATIONS AND CONCLUSIONS

Examination of the data in the table cells (Figure V,5 A - I p. 100-104:) provides a useful measure of the performance of the retrospective searching or information retrieval subsystem in the Ciba-Geigy library, and, in addition to the statistical evaluation, some interesting deductions can be made from the results.

Inspection of the values in the upper right hand cells of tables A, B, C and D of Figure V,6 p. 100 - 102) reveals values close to the ideal 1,0. The indication is, therefore, that the current retrospective searching techniques are effective. The values in the lower left hand cells are relatively low, which is an unexpected result, as these should ideally also tend to a value of 1,0. A broad interpretation of this result is that, given the wrong input, for a significant percentage of the inquiries, the correct output is, nevertheless, achieved. A possible explanation for this phenomenon may lie in the fact that, being a special library with a circumscribed stock and subject matter, the staff is able to acquire a more intimate knowledge of the library contents, thus enabling them to

exercise their own insight in the various stages of an inquiry. This familiarity with subject matter is a feature of the special library and, similarly, in-depth subject knowledge may assist information retrieval performance in other libraries with subject specialist staff e.g. university libraries.

Since the probabilities in the horizontal rows of the tables (Figure V,6 A - D) must add up to 1,0, the values in the upper left-and lower right hand cells are very low and relatively high respectively. These values require no further comment as they simply confirm the observations made earlier.

With reference to section H of the statistical measures (Figure V,6 p.103), it is seen that the proportion of relevant documents retrieved (recall) has a value of 0,81, which is considered to be very high. This is consistent with the high levels of effectiveness found, and with the postulated advantages pertaining to the special library i.e. limited subject matter and knowledgeable staff.

The proportion of non-relevant documents retrieved is, at 0,26, relatively high. The indication is perhaps that too many documents are retrieved for each inquiry. However, this is coupled with a high proportion of relevant documents retrieved, and the acceptable level of information retrieval in this pharmaceutical industry library system is possibly due to the demands for accurate timeous information because of the serious consequences of delay in the dissemination of this information in the actual therapeutic situation.

CHAPTER SIX

EVALUATION OF INDEXING LANGUAGE

6.1 MEASURES OF INDEXING EFFECTIVENESS

6.1.1 INTRODUCTION

The assessment of the retrospective searching process in Chapter 5 has provided an indication of indexing performance in relation to both indexes in use in the library system (see Figure V,6 C - Errors due to indexing p.101). In addition, these indexes were evaluated in Chapter 4 (see Tables IV,5 and IV,6 p.86) as regards recall and precision ratios. The purpose of this chapter is, therefore, to study both indexes in more detail from a comparative point of view in order to pinpoint deficiencies.

The method of analysis used is that described by King and Bryant ^{6.1} and is described in the next paragraph. Subsequent sections of this chapter describe the data collection, analysis of the data and the deductions made from the analysis.

6.2 THEORY

6.2.1 GENERAL

The objective of the microevaluation of indexing language in the two indexes in use in the library system of Ciba-Geigy, South Africa is to establish how well these indexes contribute to the retrieval of information from the system. It is necessary, firstly, to examine how

6.1 Ibid. p. 131 - 142.

well this indexing activity is performed in both indexes in use in the library. Secondly, it is necessary to establish how the indexing is done. In both indexes in use in the company library, the indexing is done by human beings as a result of analytical thought processes. The result of the indexing process is the assignment of indexing terms to documents for the purpose of identifying those documents predominantly by means of the subject content of the documents. Consequently, it is essential, for evaluation purposes, to test how well these indexing terms indicate or express the subject matter of each document.

6.2.2 METHODS

In order to test the accuracy of the assignment of indexing terms to documents in the collection, an indexer, familiar with the system, its users and objectives, has undertaken a comparative study of indexing consistency and accuracy, where each document has been indexed independently in both indexes in use in the library system:- viz. the Basle-produced Computer-printed index and the South African Catalogue card index.

6.2.3 SELECTION OF DOCUMENTS

Samples for measurement of the consistency and accuracy of indexing procedures and for comparison were drawn from recently indexed documents, i.e. documents indexed during January to October 1978 in the South African company's library, which were also indexed and provided by the Basle Documentation Centre on microfiche. These totalled 78 items about which comparative data were available. Figure VI.1, a Basle Documentation Centre bibliography card, shows all indexing terms assigned to that document separated by the prefix 601 - 615 and 701 - 703. (see p. 108).

Voltaren 1050 En	0218E	0218E
101 Marit, L. 102 Charoen, C. 103 Dhanayabhak, K. 104 Sukit, L. 105 Thamrongrat, K.	301 A short-term multi-centre comparison of 302 diclofenac sodium and indomethacin in 303 osteoarthritis and extra-articular rheumatism	801 40 patients each drug (mean age 45.3 years) with 802 osteoarthritis, spondylosis, non-articular 803 rheumatism and cervical syndrome were treated in 804 a multicentre double-blind study with Voltaren 805 (25 mg 3 times daily) or indomethacin. Four 806 patients were also suffering from other diseases 807 (pulmonary tuberculosis, uterine myoma, diabetes 808 mellitus and senile osteoporosis). Voltaren 809 abolished the assessable symptoms, improved 18 810 cases but was ineffective in 9. 7 patients on 811 Voltaren experienced moderate gastro-intestinal 812 reactions (side effects occurred in 11 patients 813 on indomethacin). Voltaren was discontinued in 2 814 patients. The therapeutic effect of Voltaren and 815 indomethacin was equipotent, but Voltaren was 816 better tolerated. (See also Voltaren 171).
501 IMed	501 Patient/RSP/T.B./lung 502 Patient/GEN/tumour/uterus 503 Patient/MET/Diabetes mellitus 504 Patient/MOT/osteoporosis/senile 505 Admin./tab. 506 Response 507 Tolerance 508 Effect/analgesic 509 Comparison/double-blind 510 Comparison/indomethacin 511 IND/MOT/osteoarthritis/spine 512 IND/MOT/osteoarthritis 513 IND/MOT/rheumatism/non-articular 514 IND/MOT/cervical syndrome 515 Contra-ind./review	901 0040 cases, clin., tables, lit. 8, double-blind
701 35/incidence 702 35/stomach/complaints 703 35/intestine/complaints		Ba/mb 9/9/77/VII

Figure VI.1 Basic Documentation Centre Bibliography Card,
Voltaren 218

An indexer in the local company library indexed the documents as part of the normal workload, unaware that these documents would form

part of a test to compare indexing consistency and accuracy between indexers from the two centres.

6.2.4 EXPERIMENTAL DESIGN

For the purpose of this study, it has been presumed that the indexers in the system, both local and Swiss, are university graduates, as well as experienced indexers. As the index language in both centres is English, this does not introduce a variable. In order to examine comparative indexing, it is important to note errors of omission, in other words, to analyse the assignment of indexing terms. (see Table VI,1 below).

<u>Term</u>	<u>Assigned</u>					
	<u>Correctly</u>		<u>Incorrectly</u>		<u>Omitted</u>	
	No.	%	No.	%	No.	%

Table VI,1 Breakdown of Indexing Terms for Each Document

Omissions i.e. underindexing (column 3) results in missed documents, while overindexing, viz. column 2, usually results in false retrievals if these terms are used for the search.

A term by term comparison has been undertaken on a test set of documents indexed by indexers in Ciba-Geigy, Basle and South Africa and summarized as in Table VI,2 overleaf.

First Indexer	Second Indexer		Totals
	Not assigned	Assigned	
Not assigned	N_{00}	N_{01}	$N_{0.}$
Assigned	N_{10}	N_{11}	$N_{1.}$
Totals	$N_{.0}$	$N_{.1}$	$N_{..}$

Table VI,2 Assignment of Indexing Terms by Neither, Both, or One of Two Independent Indexers.

Consistency will be measured by the formula

$$C = N_{11} / (N_{11} + N_{01} + N_{10}.)$$

Substitution in the above 2 x 2 table of an indexing judge for the second indexer allows the derivation of a measure of indexing accuracy as follows:-

Indexer	Correct indexing		Totals
	Not assigned	Assigned	
Not assigned	M_{00}	M_{01}	$M_{0.}$
Assigned	M_{10}	M_{11}	$M_{1.}$
Totals	$M_{.0}$	$M_{.1}$	$M_{..}$

Table VI,3 Assignment of Indexing Terms - Indexing Judge

This table can be described in terms of conditional probabilities as follows:

<u>Correct indexing</u>		
<u>Indexer</u>	<u>Not assigned</u>	<u>Assigned</u>
Not assigned	P_0	P_1
Assigned	P_2	P_3
Total	$P_0 + P_2 = 1$	$P_1 + P_3 = 1$

Table VI,4

Therefore, indexing accuracy may be described as

$$P_3 = M_{11} / M_{.1}$$

6.3 DATA COLLECTION

In order to implement a comparative evaluation of the two indexes presently in use in the library system in terms of the theory outlined above, it was necessary to establish whether the 78 documents indexed by both centres would constitute a sufficiently large sample for the meaningful application of the data in terms of the theory outlined above. It was found that these documents provided as many as 1188 different indexing terms for comparison - a sufficiently large sample. These terms were then compared for consistency and summarized as in Table VI,2, p. 110. Thereafter, the same terms were examined by an independent judge who grouped them according to the 2 x 2 table, Table VI,3 p. 110, in order to establish a measure of indexing accuracy from the sample.

6.4 ANALYSIS

The first indexer in Table VI,5 overleaf corresponds to the

company's indexer in Basle, while the second indexer is the local indexer.

First indexer	Second indexer		Totals
	Not assigned	Assigned	
Not assigned	N_{00} 301	N_{01} 89	$N_{0.}$ 390
Assigned	N_{10} 946	N_{11} 319	$N_{1.}$ 1265
Totals	$N_{.0}$ 1247	$N_{.1}$ 408	$N_{..}$ 1655

Table VI,5 Summary of the Assignment of Indexing Terms by Neither, Both, or One of Two Independent Indexers.

Consistency (C) therefore equals $N_{11} / (N_{11} + N_{01} + N_{10})$ ----- (1)

$$C = \frac{319}{319 + 89 + 946} = \frac{319}{1354} = 0,24$$

Substitution of the indexing judge for the second indexer provided the following results.

Indexer	Correct indexing		Totals
	Not assigned	Assigned	
Not assigned	M_{00} 0	M_{01} 44	$M_{0.}$ 44
Assigned	M_{10} 299	M_{11} 317	$M_{1.}$ 616
Totals	$M_{.0}$ 299	$M_{.1}$ 361	$M_{..}$ 660

Table VI,6 Assignment of Indexing Terms - Indexing Judge

		Correct indexing	
Indexer		Not assigned	Assigned
Not assigned	P_0	0	P_1 0,47
Assigned	P_2	1,0	P_3 0,53
Total	$P_0 + P_2 =$	1	$P_1 + P_3 =$ 1

Table VI,7

A measure of indexing accuracy is, therefore, found by the equation

$$P_3 = \frac{M_{11}}{M_{.1}} \quad (2)$$

$$P_3 = \frac{361}{678} \\ = 0,53$$

6.5 OBSERVATIONS AND CONCLUSIONS

Examination of the consistency and accuracy data provides a useful measure of the comparative performance of the two indexes and some interesting deductions can be made from the results.

If it can be assumed that most indexing errors are errors of omission ^{6.2}, then a strong correlation is to be expected between $(P_3)^2$ and C as defined in equations (1) and (2). This is the case with these two indexes as $(P_3)^2 = 0,28$ is highly correlated with equation (1) $C = 0,24$.

6.2 Ibid. p. 139

6.6 CIBA-GEIGY LTD: COMPARATIVE INDEXING

6.6.1 INTRODUCTION

The preceding section has illustrated the basic weakness of the Basle indexing system - overindexing. The assignment of too many terms leads to false retrievals, but, overall, the basic subject content of the documents under discussion was expressed by the use of the same correct terms: viz. Basle 299 correct and South Africa 317 correct.

It has been found in Chapter 5 (see p. 104) that the overall retrospective searching techniques are effective and that relevant documents are retrieved by the system, given that the encoded request was relevant in 95% of the queries in the sample. It remains therefore, to examine both these indexes in greater detail.

Subject indexing is the key operation in the input of documents to the information retrieval system. The index itself is the key tool in the retrieval of documents or document surrogates which are relevant to the inquirer's request. By the process of indexing, documents are subdivided according to certain characteristics - usually those of authorship and subject content. Each document will be assigned an author heading and subject heading/s. In theory, there is no restriction to the number of subject headings which may be assigned to a particular document but, in practice, indexing systems usually have limitations imposed e.g. more subject entries will be possible in a computer stored system than in a card catalogue because of storage space requirements. Each index entry assigned to a document offers an access point to the document and multiple entries allow multiple access points. There are,

however, certain limitations to the mere creation of numerous entries for each document. These factors will be dealt with later.

The index is a mechanism which brings together documents of like authorship and/or subject content (and separates documents of unlike authorship and/or subject content) even if these documents have different locations or are temporarily removed from the collection on loan. By assigning a number of different subject headings to a document, the physical restriction of assigning one shelf location to that document is overcome. The purpose of the subject indexing of documents is, therefore, to bring together related subjects and to separate or divide documents according to their differences. However, the chief purpose of the subject index must be seen in relation to the information user, and that is to provide the user himself with those documents grouped according to like characteristics which may be required in response to the user's expressed information need.

For the purposes of this study, subject indexing is emphasized i.e. the classification of documents according to their subject content. In subject indexing, the first step is to make an index description of the document to be indexed i.e. to analyse its subject content. This analysis is an analysis of the subject/subjects expressed in the document and does not demand a knowledge of the subject as an area of knowledge - as a discipline. "Subject analysis thus selects and names subject concepts as they are expressed, and sought, in documents" ^{6.3} This statement emphasizes the interplay of searcher, indexer and index.

6.3 BROWN, A.G.: Introduction to subject indexing: a programmed text. v.1.
London, Clive Bingley, (1976). p. 39(38).

"Indexing is part of a complex of operations carried out in a library and cannot be properly understood except in its perspective, that is, against the background of the overall structure and functioning of the library organization." ^{6.4} The survey carried out by Bhattacharyya found that, in his sample of medical libraries in the U.K., none of these libraries used nonconventional indexing systems such as post-co-ordinate indexing systems and he discovered that libraries in the life sciences subject field are more dependent upon "effective exploitation of the information already held." Of the company indexes discussed, all are pre-co-ordinate indexes.

Time spent at the input stage, the indexing, may save on output time and effort during the searching and retrieval process. One means of time-saving at the input stage is vocabulary control - in which some control is imposed on the indexing language in order to encode similar concepts uniformly. The size of the vocabulary is important insofar as the larger the number of index terms in the vocabulary, the greater the specificity and precision capabilities of the system will be. The vocabulary specificity must be directly related to the specificity of requests made to the system. This factor indicates the necessity to analyse requests, in order to monitor the need to modify indexing language in any system over prolonged periods of operation. It will prove uneconomical to expend a great deal of time developing a more specific indexing vocabulary than the level of specificity required by the user demands made upon the system.

The development of a comprehensive index entry vocabulary will improve recall and precision from the system, as well as shortening the

6.4 BHATTACHARYYA, K.: op.cit. p. 215.

indexing process by decreasing the intellectual effort involved in the decision-making process of assigning indexing terms. The decisions made are recorded in the authority list of indexing vocabulary; suitable entry terms are described; terms not adopted are assigned cross-references to the actual entry vocabulary and, therefore, the intellectual decision-making process of subsequent indexers is shortened. Subsequent indexers merely follow the entry language indicators. A comprehensive, readily accessible entry vocabulary leads to economies in the indexing operation as well as in the searching process. A controlled indexing vocabulary applied at the time of indexing will minimize time and costs at the searching stage, whereas free indexing with little vocabulary control will maximize effort at the output stage. The number of concepts chosen for indexing purposes reflects the degree of indexing exhaustivity and if a high degree of exhaustivity is reached i.e. all the important subject concepts are indexed, this is termed depth indexing.

Indexing systems can be divided into two groups a) those employing natural language terms arranged alphabetically and b) classified systems i.e. those arranged by notational symbols. As the documents to be indexed today contain more complex ideas and information, and are no longer restricted to a single field or discipline of knowledge, but are inter-disciplinary i.e. go beyond the bounds of a single discipline, they become more complex to index.

Documents such as the following create problems in the choice of subject headings and in the recognition of relationships, between things and processes, and conditions of things.

"Furosemide compared with hydrochlorothiazide: long-term treatment of hypertension."

"Double blind comparison clinical trial of metoprolol versus

propranolol in the long-term treatment of uncomplicated hypertensive patients."

"Interactions between environmental chemicals and drug biotransformation in man."

"Failure of 2h [^{14}C] - aminopyrine breath test to demonstrate microsomal enzyme induction by glutethimide."

"Plasma protein binding of maprotiline in geriatric patients - influence of α_1 - acid glycoprotein."

"Serum rifampicin concentration related to dose size and to the incidence of the 'flu' syndrome during intermittent rifampicin administration."

Some systems establish rules for the expression of the relationships of complex subjects. It may be possible to specify the kinds of relationship and to translate these into indexing language e.g. the hierarchical relationship between 'medicine' and 'anatomy' i.e. the genus/species relationship - medicine (genus); anatomy (species). This type of relationship must be distinguished from the relationship between a subject and its properties or operations.

Ranganathan conceived the idea of facet analysis in which every subject may be subdivided by five characteristics of division - its personality, its physical matter, its energy (operation), space, and time. i.e. P M E S T

For example, the properties of steel include its:

tensile strength
shear strength
compressive strength
ductility
temperature susceptibility

} Personality foci

The operations which may be performed on steel include:

bending	}	Energy foci
cutting		
drilling		
rolling		

A hypothetical document with the title:

'The effect of ductility in determining the minimum radii
for bending steel sections' needs to be indexed to express
both the particular property of steel - its ductility - and the opera-
tion of bending which is performed upon it.

There are two approaches to the indexing of composite subjects.
The first is to regard composite subjects as units i.e. analyse them in-
to single concepts, then select a representative term from the indexing
vocabulary for each concept, and lastly arrange these terms in order as
determined by the structure of the indexing language. This system is
called a pre-co-ordinate system. Difficulties may arise from the imposed
structure should the subject headings used as a search basis differ from
those of the system.

Secondly, complex subjects may be approached in another way, namely,
by abandoning the use of structure and by filing descriptions under each
single concept. In searching, single concepts have to be combined other-
wise it would result in precision failures. This combination of single
concepts which are then matched to the request, is known as co-ordina-
tion or more correctly as post-co-ordination.

Post-co-ordinate indexing systems facilitate information retrieval
by increasing access to subjects in documents by providing more access
points - every word chosen by subject analysis is available as an access

point. An exhaustive pre-co-ordinate index contains many subject headings but the chances of false co-ordination in searching are increased, unless links and role indicators are used to prevent confusion between nuances of meaning of a specific topic and to specify the meaning of a term relating to that topic.

e.g. cooking in oil i.e. frying
solvents for oil
the processing of oil

The classified index which has an alphabetical subject index guide to the classified sequence falls outside the scope of this work as there is no classified index in use in the company library.

6.6.2 CIBA-GEIGY SOUTH AFRICA: INDEXING

It has been pointed out that "most libraries using post-co-ordinate indexing methods have tended to generate their own lists [of subject headings], perhaps using one of the major lists as a model." ^{6.5} This was also the case in the pre-co-ordinate index in this library. Index Medicus list of medical subject headings (MeSH) was considered as the model, but was rejected for a number of reasons.

MeSH is the "standard subject heading list for the subject cataloging of books in the library" ^{6.6} [National Library of Medicine, U.S.A.] MeSH was not sufficiently detailed in its pharmacology subject headings.

6.5 FOSKETT, A.C.: The subject approach to information. 3rd ed.
London, Clive Bingley, (c1977) p. 409.

6.6 MEDICAL subject headings 1975.
Index Medicus 16(1, part 2): vii, 1975.

e.g.

Dihydralazine see under Hydralazine.

Both these drugs are Ciba-Geigy preparations. It would be confusing to find papers on both located under one preparation only. In addition, to-date, there are 1854 documents on hydralazine alone and this figure excludes books and in-company reports. MeSH does not allow sufficient subdivision under the drug to suit information requirements in the pharmaceutical industry.

The company library authority list of subject headings covers the medical, pharmacological and pharmaceutical subject fields and is continuously updated and revised.

6.6.2.1 VOCABULARY CONTROL

Entries under synonymous terms were eliminated by the use of see references and relationships between terms were indicated by means of see also references.

Because of the complexity of the subject content of many of the documents in the library's holdings, it was necessary to index as exhaustively as the manual card system would allow. As the document may deal with several themes, these may need to be divided into themes and then linked together so as to avoid confusion and irrelevant recall

e.g. "Serum rifampicin concentration related to dose size and to the incidence of the 'flu' syndrome during intermittent rifampicin administration."

Unless "'flu' syndrome", in this instance, is linked to "intermittent rifampicin administration" the system will allow the retrieval of documents dealing with influenza - (virus infection), rather than flu

syndrome, which is an adverse drug reaction occurring only with an intermittent rifampicin dosage. This would result in a number of 'false drops' or 'misses' with resultant recall and precision failures.

It is, therefore, essential that the indexing language should accommodate relationships between concepts or themes and indicate such relationships as are necessary to avoid irrelevant recall. (see p.125-134).

The use of title words for indexing is common, but is only useful for retrieval purposes if the words in the title accurately express the subject contents of the document. Frequently, title words need to be transformed to retain consistency of indexing, in the manner specified in the authority list.

e.g. "Antidepressant drug trials", "anti-epileptic trials", "anti-hypertensive drug trials". The entry term in the vocabulary is "clinical trials" which does not express the category of drug tested. Preferring the broader term, in this case, will not lessen the recall of relevant documents, although it will affect precision.

The purpose of controlling the indexing vocabulary is to ensure consistency of expression of concepts by the system indexers into language which will indicate to the searcher all possible entry points to the subject themes contained in the indexed documents in the system. The searcher is referred from those terms not in use, to those indexing language access terms which will facilitate the searching process.

Certain controls over natural language were instituted in the authority list. This involved the elimination of synonymous terms such as

bilharzia and schistosomiasis

kidney	<u>and</u>	renal
respiration	<u>and</u>	ventilatory function
liver	<u>and</u>	hepatic
Hansen's disease	<u>and</u>	leprosy
heart	<u>and</u>	cardiac
hyperactivity	<u>and</u>	hyperkinesia
hypokalaemia	<u>and</u>	potassium depletion

Truncation is another means of standardizing different word forms.

e.g.

CHILD for children, childhood

It is frequently necessary to associate two terms and in certain cases word pairs allow more retrieval precision than each term singly.

e.g.

Allograft rejection

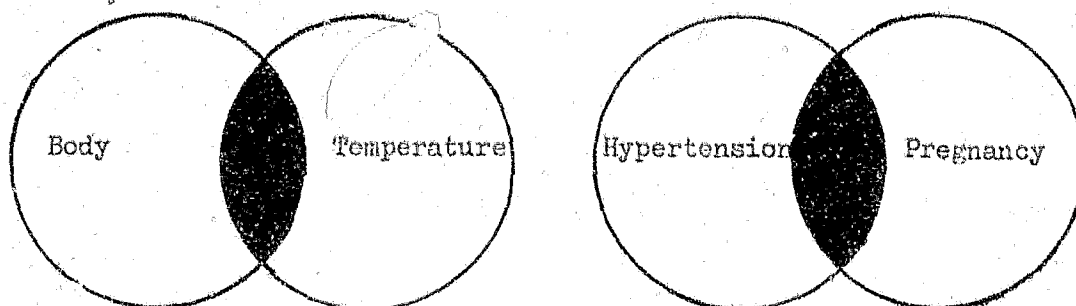
Hypertension pregnancy

Anorexia nervosa

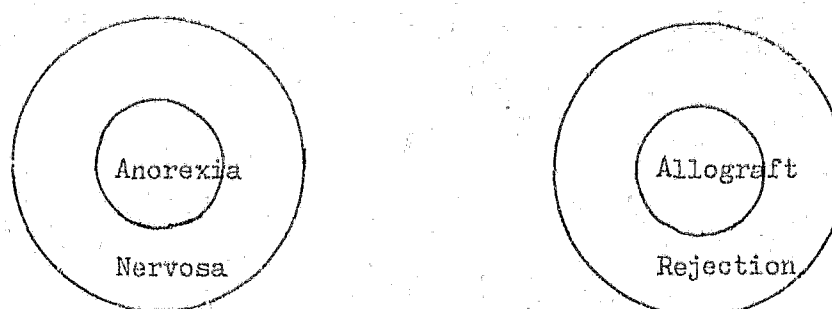
Anxiety neurosis

Body temperature

In three of the above examples there is an intersection of two classes, shown schematically as



whereas the remainder are more correctly represented as follows:



In the first instance each of these groups form a class, whereas in the second instance attributes of the concept are shown. The problem of relating concepts will be dealt with under Chapter 6.6.2.2. p. 125.

Most concepts are denoted by nouns, and certain activities may be denoted by gerunds e.g. stuttering, poisoning, weeping, learning, dreaming. Adjectival descriptions may then be added.

e.g. Epilepsy, chronic
 Epilepsy, myoclonic
 Gout, acute
 Gout, tophaceous
 Haemolysis, acute
 Hallucinations, visual
 Hypertension, severe
 Psychosis, chronic
 Purpura, thrombopenic

In this list, the singular form of the noun is usually preferred, not the plural:

Pregnancy	<u>not</u>	Pregnancies
Infection	<u>not</u>	Infections
Kidney	<u>not</u>	Kidneys

Certain exceptions do exist:-

e.g.	Malformations	<u>not</u>	Malformation
	Warts	<u>not</u>	Wart

	Veins	<u>not</u>	Vein
The spelling of various medical terms differs			
e.g.	Hyperuricaemia	and	Hyperuricemia
	Oedema	<u>and</u>	Edema
	Orthopaedics	<u>and</u>	Orthopedics
	Haematology	<u>and</u>	Hematology

It is evident that a user searching under edema (American spelling) would need to be referred to the preferred spelling, oedema.

6.6.2.2. RELATIONSHIPS

There are two kinds of relationships between terms; viz. the permanent relationship between terms denoting related subjects, such as eye, vision and ophthalmology; and the association of otherwise unrelated terms into composite subjects: e.g. the history of medicine, the pharmacology of oxprenolol, the gas chromatographic determination of hydrallazine. The former relationships are known as a semantic relationships, while the latter are syntactic relationships, which represent temporary associations of terms.

Semantic equivalence relationships have been discussed above (see Chapter 6.6.2.1. p. 123-5). These decisions on synonyms, preferred spelling, and acronyms (e.g. REM (rapid eye movement) sleep) have been incorporated into the list of subject headings. The equivalence relationship indicates that there is more than one way of expressing the same concept. The indexing language, therefore, specifies the preferred term and makes provision for users who might search under the alternative term by means of cross-references from the non-preferred term to the preferred term.

Bilharzia	<u>see</u>	Schistosomiasis
Ventilatory function	<u>see</u>	Respiration
Milk	<u>see</u>	Breast milk
Phantom limb	<u>see</u>	Pain, phantom limb
Insomnia	<u>see</u>	Sleep
Tolerance	<u>see</u>	Drug tolerance
Worms	<u>see</u>	Helminthiasis

Equivalence relationships lead from the non-preferred term to the preferred term, as distinct from hierarchical and affinitive relationships which may require cross-reference in both directions.

Eclampsia see also Pregnancy

Pregnancy see also Eclampsia

Ehrlich ascites tumor see also
Carcinogenesis

Carcinogenesis see also Ehrlich
ascites tumor

Fever see also Pyrogens
Temperature
regulation

Pyrogens see also Fever
Temperature
regulation

Temperature
regulation see also Fever
Pyrogens

Eye see also Vision

Vision see also Eye

Fractures see also Orthopaedics

Orthopaedics see also Fractures

Myocardial infarction see also
Coronary heart disease

Coronary heart disease see also
Myocardial infarction

Potassium depletion see also
Hypokalaemia

Hypokalaemia see also Potassium
depletion

Tachycardia see also Heart rate

Heart rate see also Tachycardia

Hierarchical relationships are shown in the following way:-

Sleep stages see also Narrower Term REM sleep

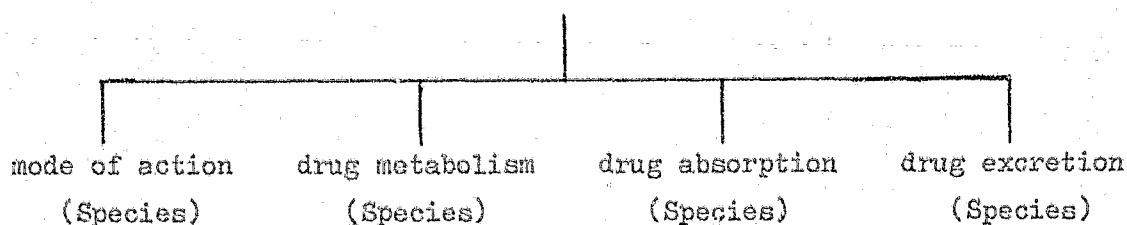
Pharmacology see also Narrower Terms Pharmacodynamics
Pharmacokinetics

Drug metabolism see also Narrower Term Half-life

i.e. the genus/species relationship is expressed.

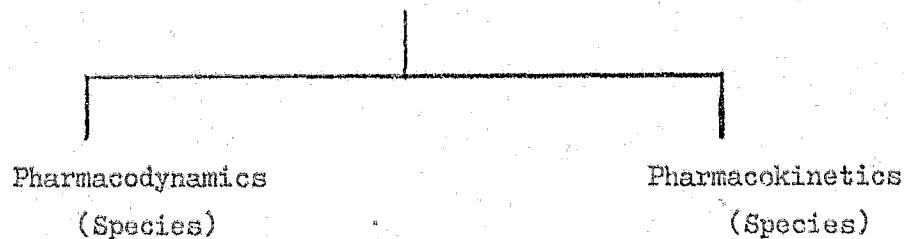
e.g.

pharmacology (Genus)



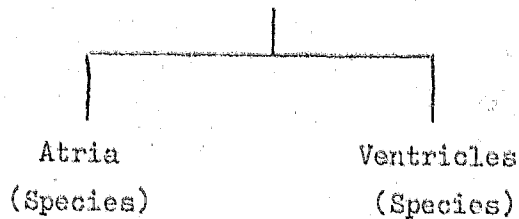
or

pharmacology (Genus)



Whole-part relationships, as in the example below, are regarded as hierarchical relationships.

Heart (Genus)



Conversely, broader terms were also indicated where it was felt that, by specifying this relationship, searches should be widened in order to increase recall.

e.g. Potassium depletion	<u>see also</u>	Broader Term Electrolyte disturbances
Rheumatoid arthritis	<u>see also</u>	Broader Term Arthritis
Osteoarthritis	<u>see also</u>	Broader Term Arthritis
Agranulocytosis	<u>see also</u>	Broader Term Blood
Antidepressant drug trials	<u>see also</u>	Broader Term Clinical trials
Neurodermatitis	<u>see also</u>	Broader Term Dermatitis
Micromelia	<u>see also</u>	Broader Term Congenital malformations

Precision is, of course, lowered if the broader term is used in searching, but, in some cases, this was found to suit the users of this library.

Affinitive and associative relationships are shown by means of
see also references:-

Encopresis	<u>see also</u>	Related Term Enuresis
Catalepsy	<u>see also</u>	Related Term Narcolepsy

Associative relationship are shown by the same means:-

Platelet aggregation	<u>see also</u>	Related Term Thrombosis
	<u>see also</u>	Related Term Allograft rejection

As accessibility is related to the network of cross-references, it is interesting to note that the Basle computer-printed index has no cross-references which, therefore, hinders user access to the index.

Inversion was used to group related terms.

e.g. Epilepsy, childhood
 Epilepsy, grand mal
 Epilepsy, masked
 Epilepsy, myoclonic
 Epilepsy, petit mal
 Epilepsy, psychomotor
 Epilepsy, temporal lobe

The above are all co-ordinate species of the genus epilepsy.

Further inverted phrases occur:-

Peritoneal adhesions	<u>Adhesions</u> , peritoneal
Alcoholic delirium	<u>Delirium</u> , alcoholic
Aplastic anaemia	<u>Anaemia</u> , aplastic
Ankylosing spondylitis	<u>Spondylitis</u> , ankylosing
Phobic anxiety	<u>Anxiety</u> , phobic
Neurogenic bladder	<u>Bladder</u> , neurogenic

The reason for these inversions is: the underlined terms are elementary subjects, with adjectival descriptors frequently in the generic relationship to the terms.

A document with the title 'The excretion of drugs in milk' was indexed as 'Drug excretion, breast milk' thereby expressing a relationship link between 'drug excretion' and 'breast milk' which is necessary unless precision is to be affected.

In the treatment of composite subjects, two kinds of relationship are important in the searching process i.e. the semantic and syntactic.

The latter is the more important in this subject field, as syntactic devices help to improve precision. A particular concept may not fit into the same category, depending on its role in the document being indexed. An example of this was the different role of 'oil' shown on p. 120. It is necessary to distinguish between oil as a product, raw material etc. These roles may be indicated by a symbol or by altering the relative position of the term. In this list, the roles are indicated in the form of subheadings.

e.g. Drug

absorption
antagonism
combination
comparison
concentration
determination
excretion
interaction
metabolism
potentiation
resistance
regulation
tolerance

Anxiety,

neurosis

Neurosis,

obsessive-compulsive

Consistency must be attained in the authority list by controlling syntactic equivalents, therefore, the combination order of composite subjects is specified in order to ensure that the composite subject is always indexed consistently regardless of how it was expressed in natural language by the author. The user undertaking a search needs to know whether he has found all similar documents indexed consistently.

Consistency is a key feature of indexing language which affects recall and precision. Combination order may create problems in the pre-co-ordinate index, whereas in the post-co-ordinate index these combinations occur at the moment of searching and there is no problem that the combination order may not suit the user as in the pre-co-ordinate index.

The other problem in combining concepts in the pre-co-ordinate index is that this creates access via only one term in the fixed combination order filing of terms. In other words, the network of cross-references from other terms in the combination should cover the subsequently combined terms, otherwise they will not be found directly in a search. This is the fault of the Basle computer-printed index and will be discussed in the next section.

It is, however, rare to find a combination order to suit all users. This problem has been partly overcome in this indexing system by providing as many headings as are thought necessary and by allowing the linking of the company's drugs to the headings in a separate sequence in the card catalogue.

e.g. A document with the title

"Interpersonal problems in narcolepsy" was indexed in the two company indexes as follows:

(see Table VI,8 overleaf)

Ciba-Geigy Basle:- subject headings	Ciba-Geigy South Africa:- subject headings
Generic name	Narcolepsy
Comparison, methylphenidate	Tofranil - narcolepsy
IND/PSY/narcolepsy	Ritalin - narcolepsy
	Drug comparison, Ritalin
	Tofranil - drug comparison, Ritalin

Table VI,8 Comparative Indexing of the Same Document

It is evident that more combinations occur in the South African company index. In addition, the combination order 'IND/PSY/narcolepsy' indicates an access point to narcolepsy which is obscure as there is no cross-reference from 'narcolepsy' to 'IND/PSY/narcolepsy'. In the company's Basle-originated index there are three access points, as opposed to five in the local company library index. The provision of fewer access keys renders the former index less easy to use and more liable to recall and/or precision failures. The fact that the former index is computerized would seem to indicate that more access points per document would be facilitated, yet, in this case, the same document accessioned three months prior to its receipt on microfilm from the Basle Documentation Centre was given two additional headings in the manual system.

Another document entitled 'The drug treatment of phantom limb pain' might be thought of as Pain: drug treatment

Pain, phantom limb

but has been indexed under specific drugs:-

Anafranil - pain, phantom limb

Tofranil - pain, phantom limb

Pain, phantom limb

with a cross-reference

Phantom limb pain see Pain, phantom limb

A more complicated composite subject such as "gas chromatographic determination of the secondary amine alprenolol as its trifluoroacetyl derivative, at nanogram levels in biological fluids" needs to be searched for under a combination of terms.

'gas chromatography, alprenolol, metabolite, determination (drug determination), biological fluids' where gas chromatography is the method used in the detection of the metabolite of alprenolol in biological fluids. Certain hidden relationships need to be expressed by the indexer i.e. the fact that the trifluoroacetyl derivative is a metabolite of alprenolol i.e. whole-part relationship.

In this case the access points are:-

alprenolol - drug determination

alprenolol - chromatography

The particular chromatographic method is unimportant, as is the metabolite of the drug, in the context of the users' requirements from the system. The omission of some possible access points may lead to recall and precision failures but, in practice, the non-use of chains of terms does not seem to affect the operating efficiency of the index (see Chapter 5 p. 101).

An example of the degree of indexing exhaustivity attained in the indexing of composite subjects is seen in the indexing of the following document:

"Antiplatelet drugs - a new pharmacologic approach to the prevention of thrombosis", in which title words are supplemented by terms or concepts taken from the text of the document. This is known as enriched title indexing and improves information retrieval from indexed documents.

The index entry terms were:-

Thrombosis

Platelet aggregation

Drug comparison, aspirin

Drug comparison, dipyridamole

Drug comparison, clofibrate

Drug comparison, dextran

Amaurosis fugax

Cerebral ischaemia

Where a document mentions or deals with a company product, Anturan, this product is assigned to a separate sequence in the card catalogue with a link to the relevant preferred index term e.g. Anturan - thrombosis.

6.6.2.3. CONCLUSION

Because of the large number of documents stored in the library, which the user will not have read and which are unfamiliar, the indexer must consider the potential user when analysing the documents for indexing. The indexer, therefore, indexes according to the actual users' backgrounds and thus chooses indexing terms which will be more likely to allow access to the documents in the store than if the indexer were to index according to the text of the documents only. There is, therefore, a reasonable relationship between the anticipated search terminology and the subject content of the documents. The

judgement of meaning and the significance of words plays an important part in the recall function of the system and the authority list of terms guides the user to the indexing terms in use in the system. If the indexer comes into contact with the users, this facilitates the addition of terms to the authority list, which are non-significant for indexing purposes, but are used by the inquirer to verbalise his request i.e. the actual vocabulary of the user. As the indexers in this library handle most of the inquiries, this has facilitated the addition of these non-significant indexing terms to the authority list of terms.

6.6.3 CIBA-GEIGY SWITZERLAND: INDEXING

The alphabetical list of terms currently in use in Switzerland was developed by Ciba and adopted by Ciba-Geigy from 1971 onwards. It covers the medical and pharmaceutical subject fields. There were some inconsistencies in indexing vocabulary in the printed indexes dating from the early 1960's. These inconsistencies have mostly been rectified in the later annual indexes sent to group companies.

The list of indexing terms is not supplied to group companies as it is not expected that they will index documents: it is considered that their recall needs will be met by the use of the annual computer-printed indexes. This, in practice, is not always the case, as current information requirements cannot be met because of the time delay in both the receipt of the microfilmed documents and the indexes to their contents. There is, however, a current awareness bulletin, the Literature Bulletin, from the Pharma Documentation Centre, Basle, which is received weekly and helps to meet the need for current information.

The computer-printed indexes form part of a computerized information retrieval system operated from company headquarters in Basle. Current accessions of periodical articles, theses, abstracts and conference proceedings are provided to group companies on microfiche. The year's accessions of items on company pharmaceutical products are indexed and the annual indexes are provided by the middle of the following year. In other words, there is an eighteen-month delay in the receipt of a cumulated index covering the previous January's documents. The library's holdings of documents on microfiche is large in comparison with its book and periodical holdings. An effective index is, of necessity, a prerequisite for access to the information contained on microfiche.

6.6.3.1 VOCABULARY CONTROL

There are many inconsistencies in indexing vocabulary to be found, particularly when comparing the early and recent indexes. This has often presented a barrier to inquirers unfamiliar with the index.

Examples of inconsistencies in language usage in the earlier annual indexes (circa 1963 - 1965) follow,

COMPARISON WITH CHLORPROTHIXINE

COMPARISON WITH COMPLEXING AGENTS

COMPARING WITH DCI ...

COMPARISON WITH OTHER ANTIHYPERTENSIVE DRUGS,

compared with the later

COMPARISON/BRADYKININ

COMPARISON/DIET

COMPARISON/FRUCTOSE

COMPARISON/FRUSEMIDE

COMPARISON/HISTAMINE

The subject index sequence of the 1964 index shows further inconsistencies

ELECTROLYTE/K/, EFFECT ON

ELECTROLYTE/NA/, EFFECT ON

ELECTROLYTES/SERUM/, EFFECT ON

in the use of both the singular and plural forms of the noun 'electrolyte'.

Furthermore, there are antonyms such as

S.T.H. / RELEASE

which are obscure and are not cross-referenced.

Chemical formulae are used inconsistently.

e.g. EXCRETION/NA/, EFFECT ON

EXCRETION/NA/, IN THE URINE/EFFECT ON

and COMPARISON/SALT meaning sodium

COMPARISON/DIET/LOW NA

and COMB/SLOW SODIUM not SLOW NA.

The earlier indexes use:

SALIVARY GLAND

SALIVARY GLAND/STRUCTURE

SALIVA/FORMATION

while the 1977 index uses

SALIVARY GLAND

and SALIVATION.

Other inconsistencies include the use in the later indexes of the term 'patient' with links to various disease processes, which are

not regarded as helpful in information retrieval, mainly because these links serve little purpose in the context of actual inquiries. This term was not used in the early indexes.

e.g. PATIENT/MOT/AMPUTATION
 PATIENT/MET/DIABETES MELLITUS
 PATIENT/PSY/DEPRESSION (1977 Index)

More examples in inconsistencies follow. Index dates are enclosed in parentheses.

ADMIN/ORAL (1967) and ADMIN/ORAL
 ADMIN/PER OS (1977)
 DOSE-RESPONSE RELATIONSHIP (1977) and DOSIS-RESPONSE
 RELATIONSHIP (1977)
 METABOLISM/MILK (1977) and MILK/CONSUMPTION (1977)
 PSYCHE/AGGRESSIVENESS (1977) and PSYCHE/ AGGRESSIVENESS
 (1977)

Synonyms have been eliminated.

e.g. KIDNEY not Renal
 SCHISTOSOMIASIS not Bilharzia
 HEART not Cardiac
 HYPERACTIVITY not Hyperkinesia
 RESPIRATION not Ventilatory function

Spelling of medical terms has been standardized, and the Swiss index prefers the English spelling of medical terms to the American.

e.g. OEDEMA not Edema
 HAEMATOMA not Hematoma
 BEHAVIOUR not Behavior

Other inconsistencies occur in the index terminology for experimental works in animals. In one case, there is

HEART/RATE/ANIMAL EXP., while under 's' is found
 SALIVARY GLAND/ANIMAL EXP., as well as
 SPECIES/CAT
 SPECIES/DOG
 SPECIES/GUINEA-PIG
 SPECIES/GUINEA-PIG/ANIMAL EXP.
 SPECIES/INVERTEBRATES/OTHER
 SPECIES/MOUSE
 SPECIES/RABBIT
 SPECIES/RAT
 SPECIES/SHEEP, then
 SPLEEN/ANIMAL EXP.

The recall of a hypothetical document about animal experiments on heart rate would be possible, in theory, provided the access point, HEART/RATE was located, but the precision ratio would almost certainly be low as a number of documents indexed under the more obvious SPECIES access point would need to be checked and considered for relevancy. Perusal of the documents indexed under SPECIES/GUINEA-PIG and SPECIES/GUINEA-PIG/ANIMAL EXP. show that both documents described experiments in guinea-pigs only and the reason for the vocabulary difference remains obscure. Similar inconsistencies in indexing language abound from 1963 to date.

6.6.3.2 RELATIONSHIPS

The later annual indexes show a considerable improvement in depth of indexing. Not all these changes have resulted in improved accessibility e.g. the term HYPERTENSION is now expressed as IND/CIR/HYPER-

TENSION. What has occurred is the disease has been linked to the bodily system affected i.e. the circulatory system 'CIR', IND is the abbreviation for indication and its very use frequently results in access point inconsistencies. e.g. IND/KIDNEY DISEASE and KIDNEY/FUNCTION might occur in the same product sequence. In the above instance, there is no cross-reference from this concept to the second term.

Table VI,9 overleaf shows how pathology indexing terms are separated throughout a sequence by the inclusion of a relationship to the field of medicine or medical speciality which deals with each. e.g. DER - dermatology, NEU - neurology, PSY - psychiatry, CIR - circulatory system. It is evident that, as these terms are not rotated, ANGINA PECTORIS is not found in the alphabetical sequence under 'A', but is obscured under IND/CIR. Thus the user of this index becomes confused at all the possible access points and the relevance predictability of the index is lowered.

A further example of obscurity of access point in these indexes is illustrated by an inquiry in which a veterinary practitioner wished to know whether a particular drug could be used for a dog with biliary fever. The index does not list veterinary indications and all entries under IND were checked. The alternative entry point was SPECIES/DOG which yielded results. Eventually, upon perusing the actual microfiche summary a relevant document was found. This was a linkage failure in the index and indicates that the relevance predictability of the index was, in this instance, poor. (see Keen, E.M). Keen explains relevance predictability as "the intelligibility to the searcher of what he finds in an index... the end result of a search may be judged by the accuracy of the searchers' prediction that citations perceived to be relevant do prove to be so when the full document is obtained and examined. A high relevance predictability would mean that more

INDEX DC ENGLISH , KEY INDEXING TERMS

31-12-74/J

58

0944

COMB./HISTAMINE/ANIMAL EXP.
0497COMB./HYDRALAZINE
0511COMB./HYDROCHLOROTHIAZIDE
0490 0511 0525COMB./HYDROCHLOROTHIAZIDE/ESIDREX
0491COMB./ISOPRENALINE
0475 0483 0496 0520COMB./ISOPRENALINE/ANIMAL EXP.
0522COMB./K
0499COMB./METHYLDOPA
0489 0523COMB./NITROGLYCERIN
0486COMB./NS/CNS/ECT
0526COMB./NS/CNS/ECT/ANIMAL EXP.
0502COMB./NS/NERVES/STIMULATION/ANIMAL EXP.
0477COMB./NS/SYMPATHETIC/STIMULATION/ANIMAL EXP.
0485COMB./NS/VAGUS/BLOCKADE
0520COMB./NS/VAGUS/STIMULATION/ANIMAL EXP.
0485 0521COMB./ORCIPRENALINE/ALUPENT/ANIMAL EXP.
0505COMB./ORTHOSTATISM
0483COMB./PHENTOLAMINE
0496

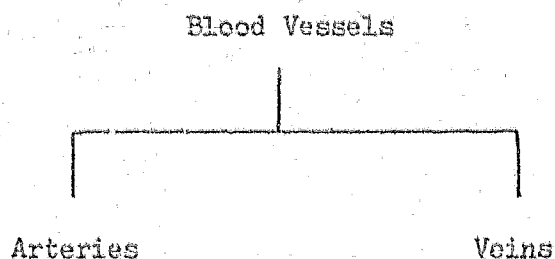
Table VI.9 1974 Cumulated Index - Product 944 (TRASICOR)

often than not a citation perceived to be relevant proves to be so: the relevant judgement can only be made on what the index provides, the index terms..." 6.7 The difficulty in determining the access point, as in this instance, indicates that this may have proved to be the reason for the students' low recall and precision ratios using this index in the experiment described in Chapter 4 (see p. 86).

There are no cross-references from the non-preferred to the preferred terms. Hierarchical relationships are, however, sometimes shown.

e.g. BLOOD PRESSURE/BLOOD VESSELS/ARTERIES
BLOOD PRESSURE/BLOOD VESSELS/VEINS

The above example shows a whole-part relationship.



Similarly, there are examples such as:-

HEART/ATRIUM
HEART/VENTRICLE
HEART/MYOCARDIUM
and EXTREMITIES/ARM
EXTREMITIES/LEG

6.7 KEEN, E. Michael: A retrieved comparison of six published indexes in the field of library and information science.
In: Unesco bulletin for libraries 30(1): 30, 1976.

and EYE/AQUEOUS HUMOR
EYE/PUPIL

Further examples of hierarchical relationships are expressed:-

PSYCHE/AGGRESSIVENESS

PSYCHE/ANXIETY

PSYCHE/ATTENTIVENESS

PSYCHE/CONCENTRATION

PSYCHE/DRIVE

PSYCHE/HOSTILITY

PSYCHE/MEMORY

PSYCHE/MOOD

PSYCHE/PERSONALITY

PSYCHE/SOCIABILITY

and TEST/COOMBS

TEST/IMMUNOLOGICAL

TEST/PSYCHOLOGICAL

as well as:

SPECIES/CAT

SPECIES/CHICKEN

SPECIES/DOG

SPECIES/FROG

SPECIES/GUINEA-PIG

SPECIES/MONKEY

SPECIES/MOUSE

SPECIES/RABBIT

SPECIES/RAT

BLOOD PRESSURE/DIASTOLIC

BLOOD PRESSURE/SYSTOLIC

It is necessary to distinguish between the measurement of blood

pressure in the various bodily positions. This relationship is not expressed, as there is merely an alphabetic arrangement of terms under 'Blood pressure'.

BLOOD PRESSURE/MEASUREMENT
 then BLOOD PRESSURE/SITTING
 BLOOD PRESSURE/STANDING
 BLOOD PRESSURE/SUPINE

This relationship ought to have been expressed as:

BLOOD PRESSURE/MEASUREMENT/SITTING etc.

Within the product sequence, terms are arranged alphabetically, and strings of terms expressing hierarchical relationships as well as associative relationships are frequently interspersed throughout the sequence.

e.g. IND/PSY/SLEEP DISORDERS under 'I'
 and SLEEP/ANIMAL EXP. under 'S'

6.6.3.3 CONCLUSION

The non-use of cross-references hinders the formulation of a logical search strategy in these indexes. Terms chosen for the search strategy frequently need to be discarded and new ones substituted until an access point is found. Thus the searcher is sometimes compelled to study the entire alphabetical sequence under one product in order to ensure that important indexing terms have not been omitted from the search strategy. Naturally, this is time consuming and inefficient and a reason for recall and precision failures. Index accessibility is therefore poor, particularly with regard to composite subjects. Combination order is inconsistent and although it is rare to find a combination order to suit all users, it is apparent that this has become a barrier to the effective use of these indexes.

6.6.4 GEIGY, UNITED KINGDOM: INDEXING

6.6.4.1 INTRODUCTION

An in-house computerized information retrieval system - in use from about 1970 - was developed to bridge the gap between the development of the Ciba-Geigy, Basle, Medical Documentation System for the Geigy range of drugs. This information system was discontinued in 1973 when the Basle Computer-printed indexes were issued to cover the Geigy product range.

The system was computerized with the main sequences being: author, 'World Health Organization International Classification of Diseases' codes, keyword, and product sequences. It was intended to provide a variety of sorted sequences not possible in a manual system.

Incorporated into the system was the ability to search for cited authors. "For example, if a certain author has done a standard work in a given sphere of a company drug then it is not unreasonable to assume that, where he is quoted in another paper, then this and following papers... [may] have some bearing on the [original] work". 6.8 This allowed searches for work on a given subject by means of bibliographical citations.

6.6.3.2 VOCABULARY CONTROL

It seems, however, from the computer sequences that no control

6.8 MD01 supplement. [unpublished]
Macclesfield, Geigy [n.d.] n.p.

was used to standardize the terminology e.g.

SIDE EFFECTS and

SIDE-EFFECTS are both used, and one finds

PRIOR THERAPY and

PRIOR TREATMENT

as well as:

INTRAVENOUS INFUSION

INTRAVENOUS INFUSIONS

also

REITERS DISEASE

REITER'S DISEASE

REITERS SYNDROME

REITER'S SYNDROME

6.6.4.3 RELATIONSHIPS

There is a long list of abbreviation keys (key letters) which allows for the expression of different kinds of relationship within the document indexed. These are index language devices which are used in association with the index terms to broaden or restrict class definition.

- D - Drugs mentioned, but without a specified link to the company product.
- DA - Additional drug to the company drug.
- DC - Drug compared with the company drug.
- DO - Other drugs.
- DP - Drug given prior to the company drug.
- E - Effect of the company drug.
- I - Indication for which company drug was given.
- O - Other indications.
- R - Reaction to company drug.
- RD - Reaction therapy.

T - Test.

TE - Test for efficacy.

TI - Test for indication (diagnosis)

TR - Test for reaction.

The ability to express these relationships is important if precision is aimed at.

The type of scientific study mentioned in the document was indicated by means of code letters, linked to both subject and product.

For example:-

D - Double blind study

G - Group

C - Case

M - Mention only

R - Review

P - Preclinical

O - Original

In addition, the printed sequences were able to serve as SDI services, alerting individuals within the company when a particular document or documents fell under their subject responsibilities. Certain key letters in a specific column on the computer print-out provided this data.

There were seven sorts on the computer which produced the following data:-

- 1) Abstracts for the Medical staff.
- 2) Abstracts for the Marketing department.
- 3) Abstracts for the Medical Documentation Service.
- 4) Product Index.
- 5) Key Letter Index.
- 6) Key Code Index.

7) Key Word/Phrase Index.

Additional information, such as the language and country of origin of documents could also be specified.

The printed, alphabetically arranged key word sequence does not use key words only, but also phrases. Inversions occur, similar to those found in the earlier Basle indexes.

e.g. RESPIRATORY SYSTEM - EFFECT ON
 RETICULAR FORMATION STIMULATION
 RNA SYNTHESIS - INHIBITION OF
 SEDATION THRESHOLD TEST
 SENSITISATION TO 5HT OF FUNDUS STRIP
 SEPTIC SHOCK. - BETA HAEMOLYTIC STREPTOCOCCI

The first, third and fifth examples express cause-effect relationships.

Genus-Species relationships are shown by means of punctuation:

e.g. SARCOIDOSIS
 SARCOIDOSIS, PULMONARY

 and TUMOUR - MOUTH
 TUMOUR - PHARYNX
 TUMOUR - LARYNX

There are inconsistencies, however, in the expression of this relationship:

 viz. TUMOUR OF VULVA

 and CANCER, PAIN
 PAIN OF CANCER

Nevertheless, the genus-species relationship is usually indicated by means of punctuation (the comma or hyphen).

Examples of the types of inquiry which may be answered using the index are as follows:

- a) Has Tegretol been used to treat herpes zoster pain?

Key letter I (indication) under Tegretol is checked and provides the following: herpes zoster

neuralgia, post hepetic

- b) Has a company antihypertensive drug been suspected of inducing diabetes?

The drug i.e. product sequence is consulted and key letter R for reactions is checked.

- c) What are the methods of analysis for Pertofran?

The product sequence is again consulted, but, this time, in conjunction with the keywords 'analysis', 'chromatography', 'electrophoresis' etc.

- d) Have Tofranil and tri-iodothyronine been given together?

In the product sequence, under Tofranil, the key letter DA is sought and checked for mention of tri-iodothyronine.

6.6.4.4 CONCLUSION

Although, in this system, considerable time will be expended at the input stage, it nevertheless results in increased precision being attained. This index allows the varied arrangement of concepts occurring in composite subjects, and, as such, improves the performance in retrieval compared with the Basle Computer-printed index. This system has almost eliminated the difficulties which arise from an imposed structural arrangement of the single concepts of composite subjects. The multiple sorts achieved in a computerized system confer this advantage.

6.7 OBSERVATIONS

From both the subjective and objective microevaluation of the two current indexes and the third discontinued index, it is apparent that there is a considerable variation of approach in the indexing of medical and pharmacological terms within this single organization. Each index possesses some favourable characteristics, not least the Geigy, U.K. index which allows the varied arrangement of single concepts occurring in composite subjects.

As regards indexing consistency, the South African index performs favourably. Its use in conjunction with the computer-printed index provides sufficiently comprehensive subject coverage of commonly indexed documents. The disadvantages of the Swiss index have been highlighted from a user access point of view.

It remains to relate index efficiency to system operation as a whole, (see Chapter 5) and it seems that the onus for the effective operation of the indexes rests on the trained information staff, without whose subject knowledge the present operating effectiveness of the system would not have been achieved.

The performance effectiveness of the system in respect of user satisfaction is dependant upon user-judged relevance of information to expressed need and the indexes are important system subcomponents. The South African index offers the user ease of access to the indexed documents, whereas the converse is true of the Basle Computer-printed index.

CHAPTER SEVEN

CONCLUSION

In previous chapters, the strong points and shortcomings of indexing and information retrieval in the library at Ciba-Geigy Ltd., South Africa have been indicated. The overall retrieval capability is better than the performance of one of the subcomponents, the Basle-produced computer-printed index as indicated in Chapters 5 and 6.

Yet the use of these facilities depends only to a small extent on their efficiency, measured from certain system characteristics. The effectiveness of the system depends on the extent to which the system characteristics correspond with the situation of the user, and on how much the potential user is willing or able to make use of the service.

As has been noted in Chapter 3, these users of this pharmaceutical information-storage and retrieval system are widely dispersed throughout Southern Africa. The question arises, therefore, of tailoring existing national information facilities to suit these needs locally. The problem, to date, has been the lack of sufficiently sophisticated drug information centres to provide quick reference results to drug and therapeutic problems encountered by the medical profession. In Britain, over the last few years, more and more drug information centres have been set up, often regionally. In South Africa, however, there have been some attempts at establishing Poison Centres on a provincial hospital level, with but one embryo 'comprehensive drug information service' attached to the Pharmacology Department of the University of Cape Town. The *raison d'être* for this latter centre appears to be the

provision of drug safety information including drug side effects, adverse drug reactions and drug interactions. It is apparent that there remains some apprehension, in this information field, that the pharmaceutical industry has some bias in the provision of information as regards the safety of its products. This fear is, to a large extent, ungrounded, however, because of the answerability of the industry to the licencing authority, the Medicines Control Council, for the efficacy and safety of all ethical preparations. This body calls for alterations to information about the drug in package inserts should use reveal new or altered tolerability and safety data.

Since the thalidomide disaster, there has been a growing public consciousness concerning drug safety. More and more governments have laid down strict drug safety and efficacy standards before the drug may be registered and marketed. Hand in hand with the growing emphasis on drug safety and efficacy has been the growth of information centres providing information on drug safety. Many of these information centres operate a 24-hour service in order to provide immediate information on drug incompatibilities, poisoning antidotes and chemotherapy. As mentioned above, similar centres exist in South Africa, but with the main emphasis on poisoning by drugs and chemicals. e.g. Addington Hospital, Durban. It appears, however, that these services will remain fragmented, because of their dependence on provincial authorities. It remains to be seen whether any national body such as the Department of Health; or the Medical Research Council could fund or organise a widespread drug information service - if indeed, this did not conflict with its objectives.

The necessity for such a service is well evidenced by the number

of inquiries reaching this one pharmaceutical company. To date, three other such companies have established information centres under trained information staff. These companies are all multinational companies and have back-up information services overseas. This factor illustrates the difficulty a local drug manufacturer would encounter, with no technical experience or information to build upon. South Africa is at present witnessing the emergence of what are called 'generic' manufacturers. These are the producers of generic so-called 'equivalent' drugs. Once the patent of an original drug has expired, these manufacturers produce the drug and market it at reduced prices. This is partly facilitated by the fact that they have had no need to fund the extensive research projects which are involved in discovering, developing and testing the drug for efficacy and safety. It is also facilitated by the fact that most such manufacturers do not provide back-up information to the medical and pharmacy professions. In fact, they are often unable to draw on the technological information available and developed by the drug innovator. It is perhaps for this reason that the existing industrial drug information centres may be called upon to provide published documents on drugs other than their own, often via inter-library loans.

The paucity of trained information staff in the field of pharmacology may partly explain why members of the medical profession have increasingly begun to utilize those existing industrial information centres. This generalization cannot be verified, but seems to be applicable to this centre.

The growth rate of this library is, to some extent, an indication of both the increase in documentation in this field and the increased user demand placed on this library from within and without the organization itself. The growth rate is currently 13% and is calculated as follows:

Annual rate of growth for:

- A Books
- B Reports
- C Periodicals

the overall growth rate is the mean of A, B and C.

- A Books = 6 %
- B Reports = 18 %
- C Periodicals = 15 %

$$\frac{6 + 18 + 15}{3} = 13 \%$$

The overwhelming problem of coping with the mass of publications in this subject field has led to the use of computerization in information processing at the Basle Documentation Centre. The service function of the local library is also dependent upon the services of highly trained information staff and more sophisticated information systems which offer the following facilities: current awareness systems; large holdings of serial publications for wide coverage of pharmacological literature; specially designed subject indexes to facilitate access to the often complicated composite subjects in documents dealing with pharmacology; comprehensive coverage of the world literature in this subject field.

It seems, therefore, that the medical and pharmacy professions' demands on this centre for information on pharmacology and therapeutics are increasing. A large proportion of the journals subscribed to by the library are held in other medical libraries, but as this centre is geared to the provision of information relevant to the inquirers'

stated information needs, it is being approached for citations in answer to actual inquiries. The emphasis on the service function of the centre distinguishes it from most local medical libraries in that a great deal of searching is undertaken by the subject specialists on behalf of the clientele.

Co-operation in this subject field has been suggested by the Medical Library Association of South Africa, which has suggested the compilation of "a list of medical, health sciences, biological sciences and related research institute libraries with notes on their resources, specialities and types of service rendered". 7.1 To date, nothing has come of this, and those libraries which are organized already co-operate through the national inter-library loan service, and contribute their serial holdings to Publications in South African Libraries - (PISAL) and book holdings to UNICAT.

In certain pharmaceutical companies in this country, pharmacists control the information function. This has meant that co-operation in the form of the submission of holdings to PISAL and UNICAT may be ruled out as the pharmacist may have no knowledge of the existence of these vehicles for co-operation. This has been verified in both the lists of PISAL and UNICAT contributors.

In limited instances, the Ciba-Geigy information staff have been approached by trained librarians for advice when these have begun setting up new information services in other pharmaceutical companies. In some cases, the existence of a collection has become apparent when the Ciba-

7.1 Medical library association of South Africa newsletter 3: 1, 1976.

Geigy library service has been approached to assist with an inquiry.

The problem regarding co-operation between libraries within the pharmaceutical industry lies in free access to material. With the fierce competition in marketing drugs, no pharmaceutical company library will provide information - even from leading medical journals - on its products to another pharmaceutical company. Hence, even if this information was published in widely read journals, it would be classed as confidential and not available to outside parties in the industry itself. Co-operation in the subject fields of interest to the pharmaceutical industry may never come about, unless placed under the auspices of some non-involved group or body, such as the Medical Research Council - which has already taken the lead in linking up with the MEDLARS network and expanding this link to include the University of the Witwatersrand Medical School Library. Costs will, however, affect the expansion of information networks such as this. This pharmaceutical company has planned the introduction of an information network for operation in the clinical pharmacology subject field. The expected operation date is set for early 1960. It is anticipated that this interactive system will increase the information demands made on the service, but that there will also be a corresponding benefit to the users within the company itself as well as the medical profession who already draw on the resources and services of this one system to a considerable extent.

BIBLIOGRAPHY

A: BOOKS

- 1 ANNAS, George J., GLANTZ, Leonard H. and KATZ, Barbara F.: Informed consent to human experimentation: the subject's dilemma.
Cambridge, (Mass.), Ballinger publishing, (1977).
- 2 ASHWORTH, W. ed.: Handbook of special librarianship and information work. 3rd ed.
London, Aslib, 1967.
- 3 BARR, Keith and LINE, Maurice, eds.: ... Essays on information and libraries; festschrift for Donald Urquhart.
London, Clive Bingley, (1975).
- 4 BECKER, Joseph and HAYES, Robert M.: Information storage and retrieval: tools, elements, theories.
New York, London, Wiley, (1973).
- 5 BERNAL, J.D.: The transmission of scientific information: a user's analysis.
International conference on scientific information, 1958. Proceedings. Washington, National academy of sciences/National research council, 1959.
- 6 BROWN, A.G.: An introduction to subject indexing. v. 1.
London, Clive Bingley, (1976).
- 7 CAMPBELL, Malcolm J. ed.: Manual of business library practice.
London, Clive Bingley, (1975).
- 8 CERLETTI, Aurelio: Drugs, the individual and society.
Basle, Pharma information, 1970.
- 9 COLLISON, Robert I.: Indexes and indexing: guide to the indexing of books, and collections of books, periodicals, music, recordings, films, and other material, with a reference section and suggestions for further reading. 4th rev. ed.
London, Ernest Benn, 1972.
- 10 DEBONS, Anthony; ed.: Information science: search for identity.
New York, Marcel Dekker, 1974.
- 11 DOE, Janet and MARSHALL, Mary Louise; eds.: Medical library association handbook of medical library practice: with a bibliography of the reference works and histories in medicine and the allied sciences. 2nd rev. ed.
Chicago, American library association, (1961).

- 12 DONOHUE, Joseph C.: Understanding scientific literature: a bibliometric approach.
Cambridge, (Mass.), MIT press, (©1975).
- 13 DRUEY, Jean: Research achievements of the Swiss pharmaceutical industry.
Basle, Pharma information, (1973).
- 14 EUGSTER, Carl: Growth planning - Geigy
In: TAYLOR, Bernard and WILLS, Gordon: eds. Long range planning for marketing and diversification. London, Bradford univ. press and Crosby Lockwood, (1971).
- 15 HAMBURG, Morris [and others]: Library planning and decision making systems.
Cambridge, (Mass.), MIT press, (©1974).
- 16 HOUGHTON, Bernard: Scientific periodicals: their historical development, characteristics and control.
London, Clive Bingley, (©1975).
- 17 INTERNATIONAL MEETING OF MEDICAL ADVISERS IN THE PHARMACEUTICAL INDUSTRY. 2nd, Florence, 1975. Rationality of drug development: proceedings.
Amsterdam, Excerpta Medica, 1976.
- 18 KEET, Allen: Information analysis and retrieval.
New York, Wiley-Becker-Hayes, (1971).
- 19 KIMBER, R.T.: Automation in libraries. 2nd ed.
Oxford, Pergamon, 1974.
- 20 KING, D.W. and BRYANT, E.C.: The evaluation of information services and products.
Washington, Information resources press, 1971.
- 21 LANCASTER, R. Wilfrid: Information retrieval systems: characteristics, testing and evaluation.
New York, Wiley, (©1968).
- 22 LEAKE, Chauncey Depew: An historical account of pharmacology to the 20th century.
Springfield, Illinois, Charles C. Thomas, (1975).
- 23 McGARRY, K.J.: Communication knowledge and the librarian.
London, Clive Bingley, (©1975).
- 24 RUBIN, Alan A. ed.; New drugs: discovery and development.
New York, Basel, Marcel Dekker, (©1978).

- 25 SCHWARTZMAN, David: Innovation in the pharmaceutical industry.
Baltimore, London, Johns Hopkins university press, (1976).
- 26 SHARP, John R.: Some fundamentals of information retrieval.
(London), Deutsch, 1968.
- 27 SHIMMON, Ross ed.; A reader in library management.
London, Clive Bingley, (1976).
- 28 SWIHART, Stanley J. and HEFLEY, Beryl F.: Computer systems in the library: a handbook for managers and designers.
Los Angeles, Melville publishing, 1977.
- 29 VICKERY, B.C.: Techniques of information retrieval.
London, Butterworths, (1971).
- 30 VICKERY, B.C.: Information systems.
London, Butterworths, (1973).
- 31 WELLSCH, Hans and WILSON, Thomas D. eds.; Subject retrieval in the seventies: new directions.
Maryland, University, 1972.
- 32 WEST GERMAN PHARMACEUTICAL MANUFACTURERS' ASSOCIATION: From molecule to medicine.
Frankfurt, the Association, 1974.

B: PERIODICALS

- 33 ARNOLD, D.V.: An industrial information service: the effects of growth on its organization and administration.
In: Aslib proceedings 16(8): 234 - 245, 1964.
- 34 BARBER, A. Stephanie, BARRACLOUGH, Elizabeth D. and GRAY, W.A.: Closing the gap between the medical researcher and the literature.
In: British medical Journal 1(5796): 368 - 370, 1972.
- 35 BARKLA, J.K.: The University of Sheffield biomedical information project.
In: Information scientist 3(1): 13 - 30, 1969.
- 36 BASILE, Victor A. and SMITH, Reginald W.: Evolving the 90% pharmaceutical library.
In: Special libraries 61(2): 81 - 86, 1970.
- 37 BHATTACHARYA, K.: Some general characteristics of special libraries in science and technology in the U.K.
In: Journal of documentation 28(3): 214 - 232, 1972.
- 38 BOON, J.A.: Inligting of raaiwerk: die inligtingsbehoefte van die drukkersbedryf aan die Witwatersrand.
In: South African libraries 41(2): 59 - 65, 1973.
- 39 BOTTLE, R.T.: A user's assessment of current awareness services.
In: Journal of documentation 21(3): 177 - 189, 1965.
- 40 BURLAND, W.L.: Role of the pharmaceutical industry in postgraduate medical education.
In: British Journal of clinical pharmacology 5 (1): 86, 1978.
- 41 BUSH, Dorcas: Problems in translation.
In: Special libraries 58(2): 173 - 178, 1967.
- 42 CHENEY, A.G.: Information dissemination in a research establishment.
In: Information scientist 5(2): 66 - 76, 1971.
- 43 CLEVERDON, C.W.: Aslib Canfield research project on the comparative efficiency of indexing systems.
In: Aslib proceedings 12: 421 - 431, 1960.
- 44 _____ : Evaluation tests of information retrieval systems.
In: Journal of documentation 26: 55 - 67, 1970.
- 45 _____ : Information and its retrieval.
In: Aslib proceedings 22: 538 - 548, 1970.

- 46 CLEVERDON, C.W.: On the inverse relationship of recall and precision.
In: Journal of documentation 28(3): 195 - 201, 1972.
- 47 _____ : User evaluation of information retrieval systems.
In: Journal of documentation 30(2): 170 - 180, 1974.
- 48 "CLINICUS": Drug information services.
In: Pharmaceutical Journal 221(5978): 28 - 29, 1978.
- 49 CLOSER liason between commercial and non commercial sources "needed".
In: Pharmaceutical Journal 219(5939): 301, 1977.
- 50 COLE, P.F.: The analysis of reference question records as a guide to the information requirements of scientists.
In: Journal of documentation 14(4): 197 - 207, 1958.
- 51 COOVER, Robert W.: User needs and their effect on information center administration.
In: Special libraries 60(7): 446 - 456, 1969.
- 52 CORBETT, L.: Controlled versus natural language: a report on the great debate.
In: Information scientist 5(3): 115 - 120, 1971.
- 53 CRAWFORD, Susan: Medical school libraries in the United States 1960 through 1975.
In: Journal of the American medical association 237(5): 464 - 468, 1977.
- 54 CURRIE, W.J.C., McCONACHIE, W.A. and TURMEK, Pat: Development of an information department in the pharmaceutical industry.
In: Pharmaceutical Journal 221(5983): 130 - 131, 1978.
- 55 DAVIDSON, Helen L.: The Lilly library - 76 years.
In: Special libraries 57(6): 391 - 394, 1966.
- 56 DAVIES, J.R.: The problem of information exchange and the complete information centre concept.
In: Information scientist 4(2): 39 - 45, 1970.
- 57 DILLENEY, Betle, H., PRESTON, Lester W. jr. and WEBB, Mary: Determining tomorrow's needs through today's requests: an automated approach to interlibrary loans.
In: Special libraries 61(5): 238 - 243, 1970.
- 58 DOLLERY, C.T. and BENNETT, P.N.: Clinical trials.
In: British Journal of clinical pharmacology 2(6): 479 - 480, 1975.
- 59 DRAGE, J.F.: User preferences in published indexes: a preliminary test.
In: Information scientist 2(3): 111 - 114, 1968.

- 46 CLEVERDON, C.W.: On the inverse relationship of recall and precision.
In: Journal of documentation 28(3): 195 - 201, 1972.
- 47 _____ : User evaluation of information retrieval systems.
In: Journal of documentation 30(2): 170 - 180, 1974.
- 48 "CLINICUS": Drug information services.
In: Pharmaceutical Journal 221(5978): 28 - 29, 1978.
- 49 CLOSER liason between commercial and non commercial sources "needed".
In: Pharmaceutical Journal 219(5939): 301, 1977.
- 50 COLE, P.F.: The analysis of reference question records as a guide to the information requirements of scientists.
In: Journal of documentation 14(4): 197 - 207, 1958.
- 51 COOVER, Robert W.: User needs and their effect on information center administration.
In: Special libraries 60(7): 446 - 456, 1969.
- 52 CORBETT, L.: Controlled versus natural language: a report on the great debate.
In: Information scientist 5(3): 115 - 120, 1971.
- 53 CRAWFORD, Susan: Medical school libraries in the United States 1960 through 1975.
In: Journal of the American medical association 237(5): 464 - 468, 1977.
- 54 CURRIE, W.J.C., McCONACHIE, W.A. and TURNER, Pat: Development of an information department in the pharmaceutical industry.
In: Pharmaceutical Journal 221(5983): 130 - 131, 1978.
- 55 DAVIDSON, Helen L.: The Lilly library - 76 years.
In: Special libraries 57(6): 391 - 394, 1966.
- 56 DAVIES, J.R.: The problem of information exchange and the complete information centre concept.
In: Information scientist 4(2): 39 - 45, 1970.
- 57 DILLEHEY, Betle, H., PRESTON, Lester W. Jr. and WEBB, Mary: Determining tomorrow's needs through today's requests: an automated approach to interlibrary loans.
In: Special libraries 61(5): 238 - 243, 1970.
- 58 DOLLERY, C.T. and BENNETT, P.N.: Clinical trials.
In: British Journal of clinical pharmacology 2(6): 479 - 480, 1975.
- 59 DRAGE, J.F.: User preferences in published indexes: a preliminary test.
In: Information scientist 2(3): 111 - 114, 1968.

- 60 DRUG information: national network of centres needed.
In: Pharmaceutical Journal 217(5895): 454 - 455, 1976.
- 61 DURRACK, David T.: The weight of medical knowledge.
In: New England Journal of medicine 298(14): 773 - 775, 1978.
- 62 ECHELMAN, Shirley: Libraries are businesses, too.
In: Special libraries 65(10 - 11): 409, 1974.
- 63 EDWARDS, Sir Ronald: The multinational pharmaceutical industry.
In: APBI news 154: 4 - 5, 1975.
- 64 EDWARDS, Tom: A comparative analysis of the major abstracting and indexing services for library and information science.
In: Unesco bulletin for libraries 30(1): 18 - 25, 1976.
- 65 ELMAN, Stanley A.: The humanization of information science.
In: Special libraries 67(9): 421 - 427, 1976.
- 66 FRANCILLON, Mary: Information retrieval: a view from the reference desk.
In: Journal of documentation 15(4): 187 - 198, 1959.
- 67 GETTING the information we need - how drug information centres can help.
In: Drug and therapeutics bulletin 16(11): 41 - 43, 1978.
- 68 GRAY, D.E.: Veterinary information: the current situation and future trends.
In: Veterinary record 98(14): 271 - 274, 1976.
- 69 HAMMEL, Robert W. and DREGER, William R.: Pharmacy student and practitioner attitudes toward the industry.
In: Medical marketing and media 13(6): 32 - 40, 1978.
- 70 HAMMER, D.P.: Reflections on the development of an automated serials system.
In: Library resources and technical services 9(2): 225 - 230, 1965.
- 71 HANSON, C.W.: Research on users needs: where is it getting us?
In: Aslib proceedings 16(2): 64 - 78, 1964.
- 72 HARLEY, A.J., LE MINOR, M. and WEIL, A.: Towards a universal search formulation language for information retrieval.
In: Journal of documentation 28(3): 202 - 213, 1972.
- 73 HOLMSTROM, J.E.: Innovations in the indexing of contributions to series of volumes.
In: Journal of documentation 31(2): 71 - 92, 1975.

- 74 ITO, Yosoji and HORIOKA, Masayoshi: Pharmacy and the pharmaceutical sciences: information activities in Japan.
In: Special libraries 61(5): 244 - 248, 1970.
- 75 JACOBS, H. Dudley: Clinical evaluation of side-effects of drugs.
In: South African medical Journal 47(30): 1365 - 1369, 1973.
- 76 JAHODA, G. and STURSA, M.L.: A comparison of keyword from title index with a single access point per document alphabetic subject index.
In: American documentation 20(4): 377 - 380, 1969.
- 77 JICK, H.: The discovery of drug-induced illness.
In: Current therapeutics 19(5): 159 - 160, 1978.
- 78 JOHNSON, Susan W. and RADER, Ronald A.: Information retrieval for the safety evaluation of cosmetic products.
In: Special libraries 69(5 - 6): 206 - 214, 1978.
- 79 JONES, Karen Sparck: A statistical interpretation of term specificity and its application in retrieval.
In: Journal of documentation 28(1): 11 - 21, 1972.
- 80 KEEN, E.M.: The Aberystwyth index languages test.
In: Journal of documentation 29(1): 1 - 35, 1973.
- 81 _____: A retrieval comparison of six published indexes in the field of library and information science.
In: Unesco bulletin for libraries 30(1): 26 - 36, 1976.
- 82 _____: Search strategy evaluation in manual and automated systems.
In: Aslib proceedings 20(1): 65 - 81, 1968.
- 83 KING, Donald W. and PALMOUR, Vernon E.:
In: Fenichel, G. ed: Changing patterns of information retrieval.
Washington, American soc. for information science, 1974. p. 7 - 33.
Quoted by MARSHALL, Doris B.:
In: Special libraries 66(11): 501, 1975.
- 84 LADENDORF, Janice: Information service evaluation: the gap between the ideal and the possible.
In: Special libraries 64(7): 273 - 279, 1973.
- 85 LANCASTER, F.W.: A study of current awareness publications in the neurosciences.
In: Journal of documentation 30(3): 255 - 273, 1974.
- 86 _____: Mechanized document control: a review of some recent research.
In: Aslib proceedings 16(4): 132 - 152, 1964.

- 74 ITO, Yosoji and HORIOKA, Masayoshi: Pharmacy and the pharmaceutical sciences: information activities in Japan.
In: Special libraries 61(5): 244 - 248, 1970.
- 75 JACOBS, H. Dudley: Clinical evaluation of side-effects of drugs.
In: South African medical Journal 47(30): 1365 - 1369, 1973.
- 76 JAHODA, G. and STURSA, M.L.: A comparison of keyword from title index with a single access point per document alphabetic subject index.
In: American documentation 20(4): 377 - 380, 1969.
- 77 JICK, H.: The discovery of drug-induced illness.
In: Current therapeutics 19(5): 159 - 160, 1978.
- 78 JOHNSON, Susan W. and RADER, Ronald A.: Information retrieval for the safety evaluation of cosmetic products.
In: Special libraries 69(5 - 6): 206 - 214, 1978.
- 79 JONES, Karen Sparck: A statistical interpretation of term specificity and its application in retrieval.
In: Journal of documentation 28(1): 11 - 21, 1972.
- 80 KEEN, E.M.: The Aberystwyth index languages test.
In: Journal of documentation 29(1): 1 - 35, 1973.
- 81 _____: A retrieval comparison of six published indexes in the field of library and information science.
In: Unesco bulletin for libraries 30(1): 26 - 36, 1976.
- 82 _____: Search strategy evaluation in manual and automated systems.
In: Aslib proceedings 20(1): 65 - 81, 1968.
- 83 KING, Donald W. and PALMOUR, Vernon E.:
In: Fenichel, C: ed: Changing patterns of information retrieval.
Washington, American soc. for information science, 1974. p. 7 - 33.
Quoted by MARSHALL, Doris B.:
In: Special libraries 66(11): 501, 1975.
- 84 LADENDORF, Janice: Information service evaluation: the gap between the ideal and the possible.
In: Special libraries 64(7): 273 - 279, 1973.
- 85 LANCASTER, F.W.: A study of current awareness publications in the neurosciences.
In: Journal of documentation 30(3): 255 - 273, 1974.
- 86 _____: Mechanized document control: a review of some recent research.
In: Aslib proceedings 16(4): 132 - 152, 1964.

- 87 LEACH, F.N.: The regional drug information service: a factor in health care?
In: British medical Journal 1(6115): 766 - 768, 1978.
- 88 LUCAS, Mary Bisset: MEDLARS or the medical literature analysis and retrieval system.
In: South African libraries 35(1): 31 - 36, 1967.
- 89 MARSHALL, Doris B.: User criteria for selection of commercial on-line computer based bibliographic services.
In: Special libraries 66(11): 501, 1975.
- 90 MARTYN, John: Evaluation of specialized information centres.
In: Information scientist 4(3): 123 - 135, 1970.
- 91 Die MEDIESE biblioteek.
In: South African medical Journal 45(36): 989 - 990, 1971.
- 92 MEYER, Roger L. [and others]: User study of current awareness and SDI at Celanese research company.
In: Journal of chemical documentation 11(1): 24 - 29, 1971.
- 93 MILLS, J.: Information retrieval a revolt against conventional systems?
In: Aslib proceedings 16(2): 48 - 63, 1964.
- 94 MURNAGHAN, Jane H.: Health services information systems in the United States today.
In: New England Journal of medicine 290(11): 603 - 610, 1974.
- 95 OATES, David: Ciba-Geigy reassesses its global future.
In: International management 28(3): 14 - 18, 1973.
- 96 OBERHOLZER, Rudi: The gestation of a new medicine.
In: Ciba-Geigy Journal 4: 11 - 16, 1972.
- 97 O'DONNELL, William S.: The vulnerable corporate special library/information centre: minimizing the risks.
In: Special libraries 67(4): 179 - 187, 1976.
- 98 ORKIN, Carole J.: Growing problem of price and prescription medicines.
In: Medical chronicle p. 6, Nov. 1976.
- 99 PADFIELD, J.M. [and others]: An interactive drug information system.
In: Pharmaceutical Journal 216(5860): 212 - 215, 1976.
- 100 PEARSON, A.W.: Fundamental problems in information transfer.
In: Aslib proceedings 23(11): 415 - 424, 1973.
- 101 PHYSIOLOGY and pharmacology today.
In: South African medical Journal 49(8): 250 - 251, 1975.

- 102 POWELL, James R.: Evaluation of Excerpta Medica on-line.
In: Special libraries 67(3): 153 - 157, 1976.
- 103 ROSENBERG, Victor: Factors affecting the preferences of industrial personnel for information gathering methods.
In: Information storage and retrieval 3: 119 - 127, 1967.
- 104 ROWE, David: Information transfer in the industrial environment - the requirements of industry.
In: Aslib proceedings 25(11): 425 - 429, 1973.
- 105 RUSSO, James B.: Profitable and nonprofitable drugs.
In: New England Journal of medicine 299(8): 156, 1978.
- 106 SAUNDERS, W.L.: Economic success: the contribution of the information scientist.
In: Information scientist 3(3): 117 - 125, 1969.
- 107 SCHULTZ, Louise: Breaking the communication barrier between searcher and literature file: an interactive guide.
In: Journal of the American society for information science 25(1): 3 - 7, 1974.
- 108 STRAIN, Paula M.: The industrial librarian as cooperator.
In: Special libraries 60(4): 209 - 214, 1969.
- 109 STRASSMAN, Paul A.: Managing the costs of information.
In: Harvard business review 54(5): 133 - 142, 1976.
- 110 SWANSON, Rowena W.: Trends in information handling in the United States.
In: Information scientist 4(3): 92, 1970.
- 111 TAGLIACOZZO, Renata: MEDLARS: its utilization in research, education and practice.
In: Methods of information in medicine 12(3): 170 - 177, 1973.
- 112 THOMSON, T.J.: The role of the pharmaceutical industry in postgraduate medical education.
In: British Journal of clinical pharmacology 5(1): 85, 1978.
- 113 TOWNLEY, Helen M.: A look at natural language retrieval systems.
In: Information scientist 5(1): 3 - 15, 1971.
- 114 VAN HALM, Johan: International cooperation or national dependence.
In: Special libraries 69(5 - 6): 201 - 205, 1978.
- 115 VICKERY, B.C.: The present state of research into the communication of information.
In: Aslib proceedings 16(2): 79 - 91, 1964.

- 116 VINKEN, Pierre J.: Future favours short communications in medical science.
In: IRCS Journal of medical science 6(4): 122 - 124, 1978.
- 117 WATSON, L.E. [and others]: Sociology and information science.
In: Journal of librarianship 5(4): 270 - 283, 1973.
- 118 WALKEY, Elizabeth M.: Communicating with management.
In: Special libraries 57: 565 - 566, 1966.
- 119 WHOSE responsibility is drug information.
In: Scrip 229: 10, 1976.
- 120 WILLIAMS, J.S.: Recent developments in microforms systems.
In: Information scientist 3(1): 31 - 38, 1969.

SUBJECT CONTENT OF BIBLIOGRAPHICAL REFERENCES

Communication 23, 107

Current awareness services 39, 85, 92

Drugs 75, 77, 98, 105, 119

Clinical trials 58

Development 8, 17, 24, 32, 96

Pharmacology 22, 101

Indexing 6, 9, 43, 52, 59, 64, 73, 76, 80, 81

Information

Analysis 18, 88

Costs 109

Dissemination 42, 115

Needs 50, 51, 57, 71

Retrieval 4, 18, 19, 21, 26, 29, 31, 44, 45, 46, 47, 66, 72, 78,
79, 83, 88, 93, 107, 113

Science 2, 3, 10, 64, 65, 81, 106, 117

Scientific 5, 12, 16, 34, 50, 60, 68, 116, 119

Services 20, 33, 35, 38, 48, 49, 56, 67, 70, 97, 110

Evaluation 20, 21, 44, 47, 80, 81, 82, 84, 89, 90, 102

Storage 4, 103, 107, 120

Systems 28, 30, 86, 94, 99

Transmission 5, 100, 104, 111, 114, 115, 116

Libraries

Management 15, 27, 28

Special 2, 3, 7, 37, 62, 97, 108, 118

Medical 11, 53, 61, 91, 94

Pharmaceutical 36, 54, 55, 67, 74, 81

Pharmaceutical Industry 13, 14, 25, 40, 63, 69, 95, 112

Research 1, 8, 13, 24, 25, 32, 96

APPENDIX 1I N D E K S B E R E K E N I N GNAAM:OUDERDOM:ADRES:TYD: $\frac{1}{2}$ -uur per vraag.

NAAM : _____

Bladsy 1. 170

COMPUTER INDEX

VRAAG 1 : Has AMBILHAR (niridazole) ever been used in the treatment of amoebiasis ?

Is AMBILHAR (niridazole) al ooit in die behandeling van amebiase gebruik ?

TYD:

Begin:

Eindig:

RELEVANTE PUBLIKASIES:

NAAM:

Bladsy 2.

171

MOONTLIKE RELEVANTE PUBLIKASIES:

OPMERKINGS:

NAAM :

Bladsy 3.

172

CARD INDEX/KAART INDEKS

VRAAG 2 : Is TANDERIL (oxyphenbutazone) indicated in the treatment of tuberculous meningitis ?

Word TANDERIL (oxyphenbutazone) gebruik in die behandeling van "tuberculous meningitis" ?

TYD:

Begin:

Eindig:

RELEVANTE PUBLIKASIES:

NAAM:

Bladsy 4.

173

MOONTLIKE RELEVANTE PUBLIKASIES:

OPMERKINGS:

NAAM :

Bladsy 5. 174

INDEX MEDICUS

VRAAG 3 : Can TRASICOR (oxprenolol) be used in the treatment of bradycardia ?

Kan TRASICOR (oxprenolol) gebruik word by die behandeling van bradikardie ?

TYD:

Begin:

Eindig:

RELEVANTE PUBLIKASIES:

NAAM:

Bladsy 6.

175

MOONTLIKE RELEVANTE PUBLIKASIES:

OPMERKINGS:

Author De Wet D E

Name of thesis A Pharmaceutical industrial medical library: indexing and information retrieval 1978

PUBLISHER:

University of the Witwatersrand, Johannesburg

©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.