

THE IDENTIFICATION OF WOOD SPECIES WITH THE AID OF THE HOLLERITH SYSTEM. ¹⁾

by

W. W. V A R O S S I E A U,

(Timber Section of the Central Institute for Testing Materials, T.N.O., Delft).

TABLE OF CONTENTS.

	Page
I. Introduction	229
II. Comparison of the English and the Dutch universal method of identification	231
A. Procedure of the two systems	231
B. Comparison of advantages and drawbacks	232
III. Design of a universal identification key according to the Hollerith system.	234
A. Some general remarks on the procedure of the Hollerith system	234
B. Lay out of features	236
C. Comparison with other systems	238
IV. Application of the Universal Decimal Classification for the indication of botanical and geographical features.	238
A. Some general remarks on the U.D.C.	238
B. Application in the Hollerith identification scheme	239
V. Summary	240
Literature	241

I. INTRODUCTION.

The last few years investigations in the field of wood anatomy have been strongly directed towards finding a universal system of identifying wood species. The problem has, however, not yet been solved definitively.

What is to be understood by a universal method and in what respects it differs from other systems, may be summarized as follows:

- 1) A universal system of identification allows of increasing the number of wood species included indefinitely.
- 2) This implies that the key does not include wood species from only one particular geographical area, but samples from all continents.

¹⁾ This paper was presented at the meeting of the International Association of Wood Anatomists at Oxford in July 1947.

3) The system should make it possible for characteristics and wood species to be simply arranged according to different points of view.

A likely arrangement is one by:

a. Structural features.

It should be possible to select species with e.g. cribriform perforation plates, included phloem and similar structural features.

b. Physical and chemical features.

Specific gravity, hardness, colour, and the presence of certain substances such as lapachol and flavone might be included.

c. Systematic units.

Representatives of an order, family, genus and species should be easy to collect from the total number of samples described.

d. Geographical distribution.

In many cases it is desirable to have a survey of the wood species originating from a particular continent or from a particular country, in so far as they occur in the system.

4) It should be possible for the identifications to be made with a pocket lens as well as with a microscope.

It will be consented that a system of identification which meets the above demands should preferably make use of loose cards, each card bearing the features of one wood species.

At the moment two types of card indexes are in use, an English system designed by Clarke (1938) and a Dutch system designed by Pfeiffer and Várossieau (1945).

Other principles were developed by Swain (1927), Bianchi (1932) and Beversluis (1943).

The scheme of Clarke has been applied in Australia (Dadswell and Eckersley, 1941) and France (Normand 1946).

Phillips (1941) elaborated it for the identification of Coniferous woods.

The idea of a card index key has also been suggested in the field of entomology (Diakonoff 1941).

Besides the similarity of making use of loose cards, the English and Dutch identification methods for wood species differ in their procedure which, in each case shows certain advantages and certain drawbacks. These will be dealt with under II.

The object of this paper is to present a procedure of the Dutch system which combines the advantages of both.

A similar work has been started in the U.S. Forest Products Laboratory at Madison, Wisconsin.

The author owes many thanks to the late Dr J. Ph. Pfeiffer, formerly Director of Research of the Laboratory of the N. V. de Bataafsche Petroleum Mij., Amsterdam, for his important suggestions and for the way in which he has made this investigation possible.

Many thanks are also due to Mr Swaab (Head of the Financial Dept. of the B. P. M. Laboratory te Amsterdam) and to Mr Groeneveld (representative of the International Business Machines Co., Amsterdam) for the many technical details which they gave the author about the procedure and application of the Hollerith system.

II. COMPARISON OF THE ENGLISH AND THE DUTCH UNIVERSAL METHOD OF IDENTIFICATION.

A. Procedure of the two systems.

In the Dutch scheme of identification the structural elements of woods are described according to a decimal system.

Five classes of features are each of them divided into four groups, and these again into subgroups, which are indicated by means of indices. It is the intention that the classification of wood species should give a grouping which links up as well as possible with the botanical arrangement according to a natural system.

On the classification scheme is based a method of identification which makes use of loose cards (*Fig. 1*). This enables the number of wood species

	I		II		III		IV		V	
A	b	2	c	9	b	4	a	1	a	2
B	b	4	a	?	a-c	1-9	a	1	a	1
C	b	2	b	2	b-c	2-8	a	1	c	5
D	b	4	b	2/3	b	2-4	a	1	a	2

Order: 17/77-78.
Geraniales.
Family:
Euphorbiaceae.
Subfamily:
Crotonoideae.
Tribus:
Chrozophoreae.
Genus:
Aleurites.
Species:
moluccana (L.)
Willd.

Nos. 1746. 2101.
Particulars: In some samples, the definitely arranged parenchyma may be partly or wholly regarded as indefinitely arranged.

Fig. 1 — Card for identification key according to Pfeiffer and Varossieau (1945). Actual size of card 12,5 × 7,5 cm.

to be enlarged constantly. The cards are not perforated, as is mostly the case with other systems of identification using loose cards, on the ground of the consideration that the classification originally designed by Pfeiffer was also intended for the use of private persons, such as foresters, engineers and wood merchants, for whom the use of a perforating machine and selecting box would be a drawback.

By means of a table symbol combinations representing features are converted into a group of two figures, according to which the cards are arranged. To identify a wood species, its features are stated on a card, after which that card is taken from the system which corresponds to the combination of figures which can be deduced from these features.

The English system of identification (Clarke 1938) also employs loose cards (*Fig. 2*), but these are perforated along the edges. Each hole in the card corresponds to a particular feature. These holes run on as far as the edges of the card only if the wood species described on it shows the feature in question. If a rod is pushed through a hole in the complete set of cards, only the cards showing this feature fall out on shaking.

This operation is repeated until one card is left bearing the name of the species required.

B. Comparison of advantages and drawbacks.

During a discussion which the author had with a number of English I. A. W. A. members in June 1946, the following points were put forward:

- 1) Though under the Dutch method the number of wood species included can be extended constantly by adding new cards, the same method must each time be followed to make an identification. In other words, the field covered by the system is unlimited, but the method itself is fixed. In the English method it is possible to determine for each individual case what is the quickest way of arriving at an identification; the order of the features used in identification is not fixed in this case. As against this, it is often difficult to make a choice from the features if the order in which they are to be considered has not been laid down. This has become clear in Bianchi's system, which has no fixed procedure either. In this method it proved necessary to find a solution for the problem.
- 2) Some investigators deny the possibility of combining an arrangement of the cards according to relationship of the species with an arrangement determined by the wood structure. This objection is based on the supposition — an incorrect one according to the author — that there is a botanical system which is capable of describing the relationship between species accurately. If it is agreed that this is not the case, a different classification of species based on the wood structure, in comparison with a (mainly) morphological system, might give new indications of this relationship. In actual practice this has repeatedly been found possible.
- 3) The numerical combination according to which the cards are arranged in the Dutch system is obtained by combining 3 features for the first number, and 6 features for the second number. This makes it difficult to find out how a certain feature is distributed over various systematic units.

In connection with what has been said under 2), it may be remarked that it is very important that the numbers determining the order of the cards should have been obtained by a combination of features.

For it is in the first place combinations of features which give indications about the relationship and not these features separately.

Some advantages of the Dutch system over the English are:

- 4) A card with the dimensions of the "international library size" (taken from the first international postcard size: 125 × 75 mm), in

A B C D E F G H		GROWTH RINGS		GEOGRAPHICAL REGIONS		PHYSICAL PROPERTIES		OTHER FEATURES		PARENCHYMA		VESSELS		RAYS		FIBRES, ETC.		J K L M N O P Q		
88																				
87																				
86																				
85																				
84																				
83																				
82																				
81																				
80																				
79																				
78																				
77																				
76																				
75																				
74																				
73																				
72																				
71																				
70																				
69																				
68																				
67																				
66																				
65																				
64																				
63																				
62																				
61																				
60																				
59																				
58																				
57																				
56																				
55																				
54																				
53																				
52																				
51																				
50																				
49																				
48																				
47																				
46																				
45																				
44																				
43																				
42																				
41																				
40																				
39																				
38																				
37																				
36																				
35																				
34																				
33																				
32																				
31																				
30																				
29																				
28																				
27																				
26																				
25																				
24																				
23																				

Fig. 2 — Lay-out of card for multiple entry key according to Clarke (1938). Features 1, 3 and 7 are shown notched. Actual size of card 20 X 12.5 cm.

- case symbols are used, can bear many more features than the English cards, though these have larger dimensions (200 × 125 mm).
- 5) The Dutch cards bear only those features which are shown by the wood described.
- The English cards always contain all features, but the majority of them — in each card separately — are never used.
- 6) Unlike the English system, the Dutch system classifies the features in a staggered manner. This method prevents subordinate features from being placed on the same level as those, covering a wider field.
- For convenience the differences between the English and Dutch methods are recorded in the following table.

Comparison of the English and Dutch methods of identification.

Points compared	English method	Dutch method
1. Number of features	limited, influenced by size of card	unlimited, not influenced by card size
2. Arrangement	features next to each other	features staggered
3. Symbols	features are not represented by symbols	features are represented by symbols
4. Punching	features are punched into the cards	features are not punched

The way in which the advantages of the two methods can be combined will be discussed under III.

III. DESIGN OF A UNIVERSAL IDENTIFICATION KEY ACCORDING TO THE HOLLERITH SYSTEM.

A. Some general remarks on the procedure of the Hollerith system.

1. PRINCIPLE.

The Hollerith system makes use of groups of features representing different criteria of examination. These criteria can be combined in a variety of ways, and for this reason the method is excellently suitable for use in a universal method of classification and identification.

The features of each group are represented by symbols (figures or combinations of figures), and these are punched into a card.

The card contains eighty columns (numbered 0—80), each consisting of 10 figures (0—9).

Each group of features embodies one or more of such columns.

As a general rule, only one figure may be punched per column.

If more than one figure must be employed per column, two cards

should be used. These only differ in that the one figure is punched into the first card, and the other figure of the same column into the second card.

Even if they differ from each other in some respects, one or more duplicates of cards can be made in a cheap and easy manner. This operation is quite automatic, and takes place by means of a master card.

An exception to the above rule (the use of one figure per column) is the representation of the letters of the alphabet, which result from punching two symbols per column.

Here, use is made of the eleventh and twelfth punching positions on the unprinted strip at the top of the card. The entire alphabet can thus be represented in 26 columns (9 + 9 + 8 columns) (*Fig. 3*). Letters may be placed in any column.

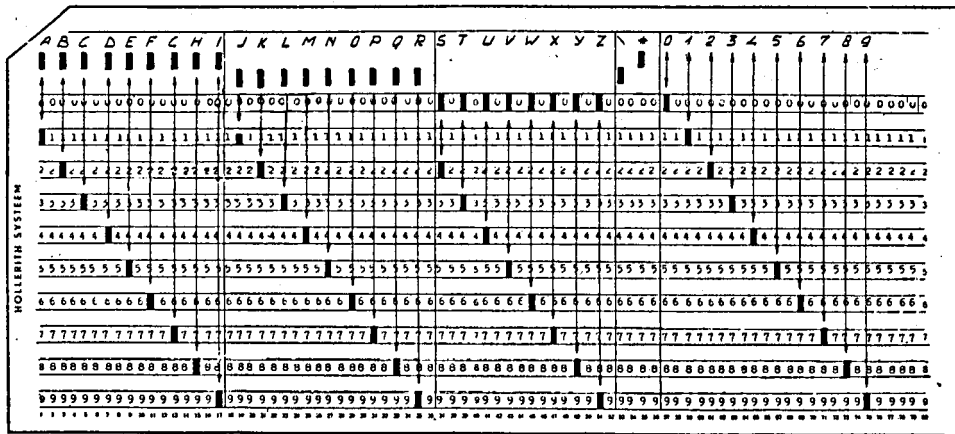


Fig. 3 — Indication of letters and numbers in a Hollerith card. Actual size of card 18,7 × 8,3 cm.

2. MACHINES.

The operation of a card index according to the Hollerith principle requires the use of three types of machines:

a. A punching machine.

Series of numbers representing definite features are punched into the cards by means of this apparatus. If necessary, particular columns which one wants to reserve, may be skipped automatically.

A punched card is checked on a second apparatus, so that errors are easily detected.

b. A sorting machine.

This machine sorts cards according to any desired column. If a pack of cards is run through the machine, a number of times, these can be sorted according to feature groups and their subdivisions.

c. A tabulating machine.

This apparatus registers the number of times which a certain feature or feature group occurs in a set of cards. The numbers are recorded automatically.

Some machines are moreover capable of "translating" feature symbols into legible letters.

Further information on the Hollerith system and other card-sorting techniques are given by Casey, Bailey and Cox (1946) and Cox, Casey and Bailey (1947). Benjamin (1947) worked out a possibility of multiple-punching with Hollerith machines.

The application of punched cards, which were originally designed for administrative purposes, to scientific problems is a.o. illustrated by Eckert (1947), King (1947) and Morgan and Frear (1947). The present writer (1947) dealt with the use in literature documentation, with special reference to wood species.

B. Lay-out of features.

The card model now designed (Fig. 4) includes botanical systematic data, distribution areas and macroscopically perceptible features as criteria

Name of the wood Fam., Genus and Spec. by U.D.C. No.	Order:	Genus:		Vernacular name:		Geo. Reg.:							
	Family:	Species:		U.D.C. No.		U.D.C. No.							
	Geograph. Regions by U.D.C. No.	Macroscopical features										Remarks.	
	I	II	III	IV	V								
	Vessels	Revs	Ray cells	Part. or	Other								
	A B C D	A B C D	A B C D	A B C D	A B C D								
0000000000	00000000												0
1111111111	11111111												1
2222222222	22222222												2
3333333333	33333333												3
4444444444	44444444												4
5555555555	55555555												5
6666666666	66666666												6
7777777777	77777777												7
8888888888	88888888												8
9999999999	99999999												9

Fig. 4 — Hollerith card, designed for the identification of wood species with the aid of a hand-lens. The lay-out of the columns 18 to 37 included is given in the publication quoted under fig. 1.

for selection. Further, a number of columns have been left open for stating native and Latin names in ordinary letters.

In connection with the operations to be performed by the tabulating machine, each of the above group of features must be numbered. The groups are indicated by the figures 1 to 4 incl. These figures only serve as an indication of a definite section of the card, and they are therefore unconnected with the figure combinations describing a certain feature in each group.

1. TAXONOMICAL DATA.

These are stated on the card as feature group 1.

2. THE DISTRIBUTION AREA.

This has been given as feature group 2.

The way in which the groups 1 and 2 are subdivided will be explained under IV.

3. MACROSCOPICALLY PERCEPTIBLE STRUCTURAL FEATURES are combined in feature group 3.

All the data relating to such features begin, therefore, with the figure 3.

The arrangement of this group is based on the above-mentioned system developed by Pfeiffer and Varossieau (1945, p. 447), the features of which, provided with a group and code number, are repeated below.

In this way the following survey is obtained:

31. Vessels.

- 311 Perforation plate
- 312 Grouping
- 313 Arrangement
- 314 Diameter and number

32. Rays.

- 321 Types and structure
- 322 Number
- 323 Width
- 324 Height

33. Parenchyma.

- 331 Paratracheal
- 332 Definitely arranged
- 333 Indefinitely arranged
- 334 Quantity

34. Particular growths.

- 341 Included phloem, oil cells
- 342 Resin canals, gum ducts and latex elements
- 343 Storied structure, radial intercellular canals, pith flecks and bast fibre formations.

35. Other characteristics.

- 351 Specific gravity
- 352 Cutting hardness and grain
- 353 Colour
- 354 Gloss and smell.

Each of the above features are divided into ten possibilities, marked 0 to 9 incl. For a description of these indices, see the 20 tables of the publication in question.

Each of the 200 features mentioned in these tables can therefore be indicated symbolically by means of three figures the figure 3 preceding as indicating the group.

Twenty columns are necessary to indicate macroscopically perceptible structural features.

4. VERNACULAR AND LATIN NAMES IN LETTERS.

In a fourth group the names of the wood species to which the card relates can be indicated by alphabetic symbols if desired.

Inclusive of spaces between words, max. 43 columns can be used for this purpose, which is more than sufficient in most cases.

As said above, modern tabulating machines are capable of indicating these names in legible letters at the top of the card. This is especially important in connection with the manifolding of cards.

In this way, the names stated, can also be reproduced automatically.

Otherwise the blank right hand side of the card may be used for stating any information desired both in handwriting or in type-writing. These data may include mechanical properties, literature references and so on.

C. Comparison with other systems.

An operation of the Dutch method of identification according to the Hollerith system allows of combining advantages of the procedures described under II.

A system of symbols is used as a result of which a great many features can be included, the mutual relations of which are shown. An additional advantage is the fact that the punching system enables a selection to be made according to any feature desired.

The procedure of the above system is limited in that use must be made of expensive machines, which will only pay if employed regularly and at full capacity.

This drawback might be met by making cards at a central office and sending duplicates to various countries at a remuneration.

For carrying out identifications or phylogenetic studies use may be made of machines which are in operation at large concerns for administrative purposes.

IV. APPLICATION OF THE UNIVERSAL DECIMAL CLASSIFICATION FOR THE INDICATION OF BOTANICAL AND GEOGRAPHICAL FEATURES.

A. Some general remarks on the U. D. C.

The U. D. C. is a classification designed as an aid in arranging data by subjects.

The term *universal* is meant to indicate that the system covers the entire field of human knowledge, and that the subjects can be grouped according to any point of view.

This is the principal point in which it differs from other systems of classification, which consider it possible to arrange subjects according to separate branches of science.

The term *decimal* denotes that the subjects are divided into 10 groups, every one of these again into 10 groups, and so forth.

The decimal character of the arrangement is not of fundamental interest; significant is the fact that subjects of unequal importance are grouped in subordination to each other.

Not all numbers 0—9 indicating the subdivision of a group are as a rule used, but some places are usually left open, which make a subsequent extension possible.

Sometimes a centesimal arrangement is desirable, i. e. if the number of equivalent subjects in a group is far higher than 10.

In some cases, as with the classification of families in the botanical system, the entire field has been divided into as many as 1000 equivalent headings (families).

The U. D. C. is derived from the system designed by Dewey in 1873. The main classification comprises the following subjects:

0. Generalities, bibliography, library science.
1. Philosophy, ethics, psychology.
2. Religion, theology.
3. Social sciences, law, government.
4. Philology.
5. Pure sciences.
6. Applied sciences, medical science, technology.
7. Art, play, sports.
8. Literature.
9. Geography, history, biography.

Since the founding of the "Institut International de Bibliographie" in 1895 five editions of the code have appeared. The third (German) edition of 1933 is the most extensive at the moment (0—5 and 62—66).

The fourth (English) edition of 1936 comprises only a few chapters from the pure sciences.

The fifth (French) edition of 1939 comprises 0, 2, 3, 61 and 65.

A number of abridged editions have appeared, inter alia, in Dutch, Swedish, Rumanian, Italian, Czech, Russian and Polish.

The extension and revision of the U. D. C. is organized by an International Committee for Universal Decimal Classification, who has now enlarged the code to include more than 90,000 headings.

The U. D. C. has so far found its widest application in literature and management documentation. The system is being used, inter alia, by the Science Library (London). Engineering Societies Library (New York) and the (still existing?) Verein deutscher Ingenieure.

B. Application in the Hollerith identification scheme.

The classification of wood species, use of which is made in characteristics group 1 of the Hollerith card, has been designed by van Heurn (1944) in collaboration with the present writer (1943). The proposals are

included in the issue of the U. D. C. now in course of preparation.

A survey of wood species might be given under three headings of the U. D. C., viz. Taxonomy (582); Forestry (634.9) and Wood Industry (674).

So-called auxiliary figures make it possible to use one and the same classification of wood species in the three chapters mentioned.

These auxiliary figures are indicated by the symbol .0.

All hardwoods start with the figure combination .031, all softwoods with .032. In this way 674.031 means hardwoods in connection with timber industry and 634.9.031 hardwoods from the forestry point of view.

The system of Engler and Prantl down through the subfamilies has been followed in the classification of softwoods and hardwoods. Three to five figures are necessary to indicate their order, family and subfamily.

The sequence of genera has been so chosen, that technically important timbers come first.

In practice most wood species are indicated by the name of the genus, in addition to which the country of origin is given.

In order to incorporate the latter commercially important feature, the distribution area has been added to the generic numbers. The figures chosen for this purpose are in accordance with those applied by the U. D. C. for indicating geographical regions. Thus 4 means Europe and the Mediterranean area, 5 East Asia, 6 Africa, 7 America (as a whole) and 9 Australia. These figures are stated only if the genus indicated occurs in more than one of the regions just mentioned.

Finally a division of species in each group is given.

Twelve to fifteen figures are necessary to indicate a wood species in the Universal Decimal Classification. The first six of them can be omitted when using Hollerith cards as they only mean "softwood" or "hardwood".

For this reason nine columns have been reserved as a maximum for stating botanical features. The method of codification mentioned makes it possible to select cards according to order, family, genus and species, which is very important in connection with studies on the distribution of characteristics in botanical units.

Geographical regions indicated in feature group 2 are given as auxiliary tables in the U. D. C. The continents are subdivided here and special subjects such as rivers and mountains are included.

This group enables a selection of cards, representing wood species from any continent or country desired.

V. SUMMARY.

The International Association of Wood Anatomists in the early years of its existence has undertaken to standardize the nomenclature used in describing woods. Later the classes of dimensions have been added thereto.

In the same way it should be possible now to standardize one or two identification methods.

Universal schemes in the first place will fit for this purpose. In the introduction it is explained which requirements should be fulfilled in such schemes.

The advantages and drawbacks of an English and a Dutch identific-

ation method are compared mutually. It is suggested, that a procedure according to the Hollerith system will allow of a synthesis of both methods mentioned, thus combining advantages and eliminating their drawbacks. The restriction in the applicability of the Hollerith scheme is determined by the fact, that complicated devices are necessary the costs of which can only be justified, if they are constantly employed at full capacity. Thus the method can only be used in a central office.

It will especially yield good results if a close international collaboration is established.

A standardized codification and centralized multiplication and distribution of cards are indispensable requirements for realising this purpose.

A short general survey is given of the Universal Decimal Classification and it is explained according to which principles wood species have been included in this scheme. The decimal codes of the U. D. C. can be used for indicating botanical and geographical data in the Hollerith identification.

In this way, the great advantage is achieved, that a literature card index on wood species can be compiled with the same figure combinations. In doing so these figures get a wider field of application than when independent classifications are made for identification and documentation purposes.

Literature.

1. BENJAMEN, K. (1947). Problems of multiple-punching with Hollerith machines. *J. Amer. Statistical Assn.* 42, 237, 46—71.
2. BEVERSLUIS, J. R. (1943). The micrographical identification of Coniferous woods. *Med. Landbouwhoogeschool te Wageningen* 47, 2, 1—39. (In Dutch).
3. BIANCHI, A. T. J. (1932). A new identification method. *Tectona (Buitenzorg, Java)* 24, 8—9, 884—893. *Ref. Trop. Woods* 29, 53, 1932. (In Dutch).
4. CASEY, R. S.; C. F. BAILEY and G. J. COX (1946). Punch card techniques and applications. *Journ. Chem. Education* 23, 10, 495—499.
5. CLARKE, S. H. (1938). A multiple entry perforated card key with special reference to the identification of hardwoods. *New Phytol.* 37, 4, 369—374.
6. COX, G. J., R. S. CASEY and F. L. BAILEY (1947). Recent developments in key-sort cards. *Journ. Chem. Education* 24, 2, 65—70.
7. DADSWELL, H. E. and A. M. ECKERSLEY (1941). The card sorting method applied to the identification of the commercial timbers of the genus *Eucalyptus*. *J. Council Sci. and Industr. Res.* 14, 4, 266—280.
8. DIAKONOFF, A. (1941). The paramount card-index applied to entomology. *Entomologische Mededelingen van Nederlands-Indië* 7, 2, 34—40. (In Dutch).
9. EKERT, W. J. (1947). Punched card techniques and their applications to scientific problems. *Journ. Chem. Education* 24, 2, 54—57 and 74.
10. HEURN, F. C. VAN (1944). Proposals for the international editions of the U. D. C., presented to the International Federation for Documentation. *Hilversum*.
11. KING, G. W. (1947). Some applications of punched-card methods in research problems in chemical-physics. *Journ. Chem. Education* 24, 2, 61—64.
12. MORGAN, J. A., and D. E. H. FREAR (1947). Coding and sorting chemical compounds by means of punched cards. *Journ. Chem. Education* 24, 2, 58—61.
13. NORMAND, D. (1946). Les clefs pour l'identification des bois et le système des fiches perforées. *L'agronomie tropicale* 1, 2—4, 162—172.
14. PFEIFFER, J. PH. and W. W. VAROSSIEAU (1945). Classification of the structural elements of the secondary wood of Dicotyledons, using decimal indices for classification and identification of wood species. *Blumea* 5, 3, 437—489.
15. PHILLIPS, E. W. J. (1941). The identification of Coniferous woods by their microscopic structure. *Linn. Soc. Journ. Bot.* 52, 343, 259—320.

16. SWAIN, E. H. F. (1927). A universal index to wood. Queensland For. Serv. For. Bull. No. 7.
17. UNIVERSELE DECIMALE CLASSIFICATIE (1944). Verkort Nederlandse Uitgave. Ned. Inst. voor Documentatie en Registratuur No. 237.
18. VAROSSIEAU, W. W. (1943). Proposals to the International Federation for Documentation on the decimal classification of wood species. Report H-43-XVII. C. I. M. O. (In Dutch).
19. — (1947). Use of the U. D. C. in selecting data with mechanical appliances. Reports 17th F. I. D. Conference, Part III, Bern.