Control
and Cybernetics
VOL. 3 (1974) No. 3/4

Data structures and their transformations*)

by

CORRADO BÖHM,
MARIANGIOLA DEZANICIANCAGLINI

Institute of Information Science
Faculty of Science University of Turin

The goal of this paper is defining a sufficiently general data structure that would allow obtaining, by means of successive classifications, many different data structures, like lists, multisets and sets. The information loss corresponding to every classification just mentioned is investigated from the point of view of the properties that are conserved. Some operations that leave such properties invariant are presented. On the other hand, correspondences between different families of structures, without information loss, i.e. one-to-one mappings, are presented.

1. Introduction

The study of simulation algorithms is closely connected with that of data structures. In fact, the simplicity of an algorithm which describes the evolution in time of a system depends essentially on the choice of the data structures representing its states. For example, the state of a lift with a memory is determined by its ascending or descending motion, by the destination of the occupants and by the requests of those who call it to go up or down. It can therefore be given as a list of three components, of which the first is the number of the current floor, preceded by plus or minus, according to whether the lift is going up or down, the second is the multiset[1]) of the non-negative floor numbers of destination of the occupants. Lastly, the third component consists of the multiset of the floor numbers corresponding to each request preceded by the sign of plus or the sign of minus, according to whether the requests are for going up or down. In this example an important role is played by the idea of the list built on a multiset. Such data structures, that is the list, the multiset and, as a special case, the set, are also useful for treating problems of combinatorial analysis.

---

*) Presented at Polish-Italian Meeting on Modern Applications of Mathematical Systems and Control Theory in Particular to Economic and Production Systems, Cracow 1972.

[1]) The multisets with n elements are the equivalence classes of lists of length n relative to the symmetric group of permutations of order bn.

Combinatorial analysis in fact often [7, 8] considers the distribution of objects among several boxes and studies how many configurations can be obtained with various types of objects and various types of boxes. Without departing from the general problem, we shall consider finite multisets of non-negative integers in non-empty boxes labeled with finite multisets of integers. This point of view is particularly useful in computer science, which talks of registers or memory positions instead of boxes, etc. The classical summarising nomenclature is:

Boxes \ Objects	Different objects	Identical objects
Different boxes	Dispositions	Compositions
Identical boxes	Classifications	Partitions

In our case, a third line is added as well as a third column to describe the general case of the multiset. In the case of objects which are all different, we shall represent them as non-negative integers. When the boxes can be distinguished, the j-th entry represents the structured contents of the j-th box. Therefore, in the first and third row we shall always have lists while in the second we shall have multisets. Below we give the detailed table indicating the nature of the structures:

Distribution scheme

Boxes	Objects		
	Different	Identical	Multiset of integers
Different	List of sets of integers Ex. $\langle\{1, 3\}, \{2\}, \{2\}, \{0, 4\}\rangle$	List of integers Ex. $\langle 4, 2, 0, 9\rangle$	List of multisets of integers Ex. $\langle[2, 0], [2, 0], [1, 1]$ $[1], [2], [2, 0]\rangle$
Identical	Multiset of sets of integers Ex. $[\{1.3\}, \{2\}, \{1,3\}$ $\{0,4\}]$	Multiset of integers Ex. $[4, 2, 0, 0, 9]$	Multiset of multisets of integers Ex. $[[2,0], [2.0], [1,1],$ $[1], [2], [2,0]]$
Multiset of integers as labels	List of multisets of sets of integers Ex. $\langle[\{1,3\}], [\{1,0\},$ $\{0,1\}], [\{0,4\}]\rangle$	List of multisets of integers Ex. $\langle[0], [4], [3, 0, 2],$ $[1,9]\rangle$	List of multisets of integers Ex. $\langle[[2,0]], [[0], [2,0],$ $[1,1]], [[0], [0,2]]\rangle$

The proposal of the present paper is to treat data structures using their "generating functions or forms". Let us define the generating form of a family of structure as the sum of linearly independent forms that corresponds one-one to the elements of the family itself (see [9] p. 130), i.e. the generating form of the family $\{X_1, \ldots, X_n\}$ is $\mathscr{F}(X'_1, \ldots, X'_n) \overset{\mathrm{df}}{=} \sum_{i=1}^{n} X'_i$, where X'_i is the form corresponding to X_i $(1 \leqslant i \leqslant n)$. The generating function corresponding to a given form is, by definition, the λ-object obtained by abstracting all free variables, occurring in the form[2]). The generating

[2]) The abstraction of a formula M relative to a variable x is denoted by λxM. Each occurrence of x in λxM is said to be bound. We say that a variable x is free in a formula M when it is not bound (see [1]).

function or form seems suitable both for representing families and for deriving properties of a family and correspondences between families.

This paper splits into two parts: the first, starting from a sufficiently general data structure, derives through successive classifications — following almost the same scheme as in [2] — the data structure most useful in computer science. In fact we can obtain, in a very natural way, from the generating function of the hereditarily finite lists (Sec. 2.1) the generating functions of data structures relative to problems of distribution (Sec. 2.2), of digraphs and functional digraphs (Sec. 2.3), of vectors of integers (Sec. 2.4), and of multisets of integers (Sec. 2.5) subject to some reasonable restrictions. Finally, Section 2.6 shows how the one-one correspondence between lists and multisets (Parikh's mapping) is reflected in their generating forms.

The second part treats the problem of the construction of a one-one correspondence between a family of data structures and the non-negative integers (*listing*); a general method is described which uses the generating form of the family to be listed starting from the hypothesis that every element of this family may be uniquely determined giving a certain number of parameters (non-negative integers) representing its properties. This method is applied to the listing of the families of vectors of non-negative integers subjected to some restrictions (Sec. 3.1.). In contrast, Section 3.2 exhibits a listing of the family of sets of non-negatixe integers as an example of a listing not using the general method described above.

2. Classification of data structures

We wish to present a sufficiently general definition of a data structure, so that, from it, through successive classifications, the most useful data structures in computer science may be produced [2].

Let us introduce some conventions to simplify the notation for formulas:

(i) $\vec{z}_{(h,n)}$ for $h<n$ denotes $z_h, z_{h+1}, \dots, z_{n-1}$;

(ii) $\vec{z}_{(n)}$ denotes $\vec{z}_{(0,n)}$;

(iii) $\vec{z}$ denotes $z_0, z_1, \dots$;

(iv) $\lambda\vec{z}_{(n)}$ denotes $\lambda z_0, \lambda z_1, \dots, \lambda z_{n-1}$;

(v) $\lambda\vec{z}$ denotes $\lambda z_0, \lambda z_1, \dots$;

(vi) $1_{(n)}$ denotes $\underbrace{1, \dots, 1}_{n}$

2.1. Hereditarily finite lists multisets and sets

Let us choose, as our fundamental structure, from which to deduce all the others, that of the hereditarily finite lists (h.f.l.) whose definitions is:

$$\text{h.f.l.} \overset{\text{df}}{=} \underbrace{\langle \text{h.f.l.} \dots \text{h.f.l.} \rangle}_{n} \quad (n \geqslant 0).$$

For the elements of this family let us agree to use the following representation′:

$$\langle\,\rangle' \overset{\mathrm{df}}{=} 1 = z_0^0,\ \langle\mu\rangle' \overset{\mathrm{df}}{=} z^{\mu'},\ (\nu \cap \langle\mu\rangle)' \overset{\mathrm{df}}{=} \nu' \cdot z_n^{\mu'} \text{ }^{3)} \tag{2.1}$$

where μ and ν are hereditarily finite lists and n is the length of ν. [For example, the representation of $\langle\langle\rangle\langle\langle\rangle\langle\rangle\rangle\rangle$ is $z_0 \cdot z_1^{z_0 \cdot z_1}$.

The *generating function of the family of hereditarily finite lists* must generate all possible monomials of the above kind. It can be calculated according to the following recurrence relation on the nesting level, n, of parentheses:

$$\mathscr{L}_0 \overset{\mathrm{df}}{=} \lambda z_0\, \mathscr{L}_0^0(z_0) \overset{\mathrm{df}}{=} \lambda z_0\, z_0^0 = \lambda z_0 1\,, \tag{2.2}$$

$$\mathscr{L}_{n+1} \overset{\mathrm{df}}{=} \left(\lambda \vec{z} \sum_{j=0}^{\infty} \mathscr{L}_{n+1}^{j}(\vec{z})\right) \overset{\mathrm{df}}{=} \lambda \vec{z} \left(\sum_{k=0}^{\infty} \left(\prod_{h=0}^{k} \left(\sum_{j=0}^{|\mathscr{L}_n|-1} z_h^{\mathscr{L}_n^j(\vec{z})}\right)\right)\right)^{4)}$$

where the order of addenda with respect to the index j is arbitrary.

The choice of theis representation is subject to opinion. Indeed, the concept of list includes the non-commutativity of the elements: this non-commutativity could be realized by means of a non-commutative product [4] or by indexing all factors of a commutative product, as here.

By definition, a multiset can be obtained from a list neglecting the order between the elements; this would mean in our representation that we ignore the indexing. Thus the representation ′ of the hereditarily finite multisets can be easily obtained from that of the hereditarily finite lists as follows:

$$[\,]' \overset{\mathrm{df}}{=} 1 = z^0,\ [\mu]' \overset{\mathrm{df}}{=} z^{\mu'},\ (\nu \oplus [\mu])' \overset{\mathrm{df}}{=} \nu' \cdot z^{\mu'} \text{ }^{5)} \tag{2.3}$$

where μ and ν are hereditarily finite multisets.

For example, the representation of the hereditarily finite multiset $[[\,][[\,][\,]]]$ is z^{1+z^2}.

As above, the *generating function* $\mathscr{M}_n$ *of the family of hereditarily finite multisets* with nesting level of parenthese, n, can be calculated according to the recurrence relation:

$$\mathscr{M}_0 \overset{\mathrm{df}}{=} \lambda z\, \mathscr{M}_0^0(z) \overset{\mathrm{df}}{=} \lambda z\, z^0 = \lambda z 1\,,$$

$$\mathscr{M}_{n+1} \overset{\mathrm{df}}{=} \lambda z \left(\sum_{j=0}^{\infty} \mathscr{M}_{n+1}^{j}(z)\right) \overset{\mathrm{df}}{=} \lambda z \left(\prod_{j=0}^{|\mathscr{M}_n|-1} \frac{1}{1 - z^{\mathscr{M}_n^j(z)}} - 1\right). \tag{2.4}$$

where the order relative to the index j is arbitrary. The subtraction of 1 is made to avoid the generation of the empty structure.

If in developing the series for $\dfrac{1}{1-z}$ in (2.4) we confine ourselves to the first two terms, we exclude those multiplicities greater than 1, and therefore generate

---

[3] $\cap$ denotes the operation of concatenation between lists.

[4] $|\mathscr{F}|$ denotes the cardinality of the family whose generating function is $\mathscr{F}$.

[5] $\oplus$ denotes the operation of multi-union between multisets.

the family of the hereditarily finite sets. The *generating function* $\mathscr{S}_n$ *of the family of the hereditarily finite sets* with nesting level of parentheses, n, is then:

$$\mathscr{S}_0 \stackrel{\text{df}}{=} \lambda z\, \mathscr{S}_0^0(z) \stackrel{\text{df}}{=} \lambda z z^0 = \lambda z 1$$

$$\mathscr{S}_{n+1} \stackrel{\text{df}}{=} \lambda z \left(\sum_{j=0}^{\infty} \mathscr{S}_{n+1}^j(z) \right) \stackrel{\text{df}}{=} \lambda z \left(\prod_{j=1}^{|\mathscr{S}_n|-1} (1 + z^{\mathscr{S}_n^j(z)}) - 1 \right). \tag{2.5}$$

We wish explicitly note that there is a one-one correspondence respectively between the families of hereditarily finite lists (multisets) and the families of ordered (non-ordered) tree structures. In this correspondence every pair of parentheses is associated with a node. The relation of immediate inclusion between parentheses pairs corresponds to the relation of "to be son" between nodes. Figure 1 shows

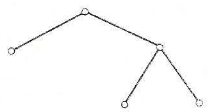

Fig. 1. A tree structure

a tree structure which corresponds, if considered ordered, to $\langle\langle\rangle\langle\langle\rangle\langle\rangle\rangle\rangle$ and to $[[\,][[\,][\,]]]$, if not.

2.2. Data structures relative to problems of distribution

In representing the non-negative integers inside our system of generating functions we have some choice. Let us agree to represent the non-negative integers by hereditarily finite multisets, with the convention that the integer k corresponds to the multisets formed by k empty multisets:

$$k \leftrightarrow [\underbrace{[\,]\ldots[\,]}_{k}] \quad \text{for every } k \geqslant 0. \tag{2.6}$$

According to this convention, for instance, the representation of 3 is $[[\,][\,][\,]]$.

In this way, every multiset with a maximum nesting level 1 corresponds to a non-negative integer, i.e. we can interpret the sum of the generating forms $\mathscr{M}_0(z)$ and $\mathscr{M}_1(z)$ as the *generating form* $\mathcal{N}(z)$ *of the non-negative integers*. From the (2.4):

$$\mathcal{N} \stackrel{\text{df}}{=} \lambda z \left(\sum_{j=0}^{\infty} n^j(z) \right) = \lambda z (\mathscr{M}_0(z) + \mathscr{M}_1(z) = \lambda z \left(\frac{1}{1 - z^{\mathscr{M}_0^0(z)}} \right) = \lambda z \left(\frac{1}{1-z} \right). \tag{2.7}$$

That is, we have achieved the well known generating function of integers in which k is represented by the monomial z^k.

An alternative representation of the non-negative integer k by the list formed by k empty lists

$$k \leftrightarrow \langle \underbrace{\langle\rangle \ldots \langle\rangle}_{k} \rangle \qquad \text{for every } k \geqslant 0$$

has been rejected because the corresponding representation of the integer k: $z_0 \cdot z_1 \cdot \ldots \cdot z_{k-1}$ was too complicated.

The *generating function* $\boldsymbol{l}$ *of the lists of non-negative integers* (*compositions* can be obtained by means of the product of copies of generating form $\boldsymbol{n}(z)$ with different indexing

$$l=\lambda\vec{z}\left(\sum_{k=0}^{\infty}\left(\prod_{h=0}^{k}n(z_h)\right)\cdot z_k\right)=\lambda\vec{z}\left(\sum_{k=0}^{\infty}\left(\prod_{h=0}^{k}\frac{1}{1-z_h}\right)\cdot z_k\right). \tag{2.8}$$

Note that, because of the multiplication by z_k, the list $\langle x_0, x_1, \ldots, x_k\rangle$ is represented by the monomial $z_0^{x_0}\cdot z_1^{x_1}\cdot\ldots\cdot z_k^{x_k+1}$. This is necessary to assure that the last exponent of the monomial ne positive, since otherwise this representation would be not unique.

Keeping this representation of integers, the generating function obtained by putting, as exponent of z in $\frac{1}{1-z}$, the monomials generated by the form $\boldsymbol{n}(z)$ becomes the *generating function* $\boldsymbol{m}$ *of the multisets of non -negative integers* (*partitions*):

$$m\stackrel{\mathrm{df}}{=}\lambda z\left(\sum_{j=0}^{\infty}m^j(z)\right)=\lambda z\prod_{j=0}^{\infty}\frac{1}{1-z^{n^j(z)}}-1\Big)=\lambda z\left(\prod_{j=0}^{\infty}\frac{1}{1-z^{z^j}}-1\right). \tag{2.9}$$

Not that in this way the multiset $[\underbrace{x_0, \ldots, x_0}_{y_0}, \ldots, \underbrace{x_k, \ldots, x_k}_{y_k}]$ corresponds to the monomial $z^{y_0 z^{x_0}+\ldots+y_k z^{x_k}}$ in which the exponent of z is the very polynomial that Knuth [5, 2, p. 551] makes corresponds to this multiset.

If, in developing the series $\frac{1}{1-z}$ in the function $\boldsymbol{m}$, we confine ourselves to the first two terms, we exclude the multiplicities greater than one, i.e., we obtain the *generating function* $\boldsymbol{s}$ *of sets of non-negative integers*

$$s\stackrel{\mathrm{df}}{=}\lambda z\left(\sum_{j=0}^{\infty}s^j(z)\right)=\lambda z\left(\prod_{j=0}^{\infty}(1+z^{z^j})-1\right). \tag{2.10}$$

The *generating function* $\boldsymbol{lm}$ *of the lists of multisets of non-negative integers*, i.e., of the structure to be found at the intersection of row 1 with column 2 in the scheme of distributions, can be obtained by means of the product of copies of the generating form $\boldsymbol{m}(z)$ with different indexing

$$lm=\lambda\vec{z}\left(\sum_{k=0}^{\infty}\left(\prod_{h=0}^{k}m(z_h)\right)\right)=\lambda\vec{z}\left(\sum_{k=0}^{\infty}\left(\prod_{h=0}^{k}\left(\prod_{r=0}^{\infty}\frac{1}{1-z_h^{z_h^r}}-1\right)\right)\right). \tag{2.11}$$

Similarly, the *generating function* $\boldsymbol{ls}$ *of the lists of non-negative integer sets* (*dispositions*) is

$$ls=\lambda\vec{z}\left(\sum_{k=0}^{\infty}\left(\prod_{h=0}^{k}s(z_h)\right)\right)=\lambda\vec{z}\left(\sum_{k=0}^{\infty}\left(\prod_{h=0}^{k}\left(\prod_{r=0}^{\infty}(1+z_h^{z_h^r})-1\right)\right)\right). \tag{2.12}$$

The *generating function* $\boldsymbol{mm}$ *of the multiset of multisets of non-negative integers* can be obtained by putting, as exponent of z in $\frac{1}{1-z}$, the monomials generated by the form $\mathscr{M}(z)$:

$$mm=\lambda\vec{z}\left(\prod_{j=0}^{\infty}\frac{1}{1-z^{\mathscr{M}^j(z)}}-1\right). \tag{2.13}$$

This structure is, in the distribution scheme, at the intersection of the second row with the third column.

Similarly, the *generating function* $\boldsymbol{ms}$ *of the classifications* (*multisets of sets of non-negative integers*) is

$$\boldsymbol{ms}=\lambda z\left(\prod_{j=0}\frac{1}{1-z^{s^j(z)}}-1\right). \tag{2.14}$$

As above, we could deduce the generating functions of the lists of multisets and sets of non-negative integers, that is, the last of the structures contained in the distribution scheme (at the intersection of the third row with the third and first columns respectively).

2.3. Digraphs

Let a digraph with k vertices be represented by a vector of length k whose elements are sets of positive integers less than $(k+1)$ or the empty set.

For example, the list $\langle\{1,2\},\{5\},\{2,6\},\{\,\},\{2\},\{\,\}\rangle$ represents the digraph of Fig. 2.

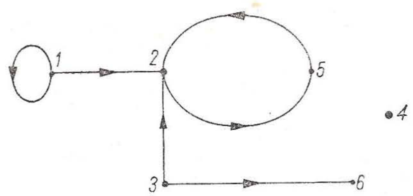

Fig. 2. Example of a digraph

The generating function $\boldsymbol{s}_{(1\div k)}$ of the sets (empty set included) of positive integers less than $(k+1)$, obtained from $\mathscr{S}$ by limiting the index j from above and below, while omitting the subtraction of 1 (the representation of the empty set), is

$$\boldsymbol{s}_{(1\div k)}=\lambda z\left(\prod_{j=1}^{k}(1+z^{z^j})\right). \tag{2.15}$$

According to the representation of the digraphs given above, the *generating function* $\boldsymbol{d}_k$ *of the digraphs with* k *vertices* can be obtained by means of a non-commutative product of k copies of the generating form $(\boldsymbol{s}_{(1\div k)}\,z)$ with different indexing:

$$\boldsymbol{d}_k=\lambda\vec{z}_{(k)}\left(\prod_{h=0}^{k-1}\boldsymbol{s}_{(1\div k)}(z_h)\right)=\lambda\vec{z}_{(k)}\left(\prod_{h=0}^{k-1}\left(\prod_{j=1}^{k}(1+z_h^{z_h^j})\right)\right). \tag{2.16}$$

From the digraphs we may pass to the functional digraphs by applying the "min" operator, defined as follows:

$$\min z_h^{z_h^r}\cdot\ldots\cdot z_h^{z_h^s}\stackrel{\mathrm{df}}{=}z_h^j,\qquad \text{where } j=\min\{r,\ldots,s\}, \tag{2.17}$$

$\min(1)\stackrel{\mathrm{df}}{=}1$, which is equivalent to replacing each set of integers in the list representing the digraph, by a single integer, the least among them, and the empty set by zero.

The functional digraphs with k vertices are in this way represented by vectors of non-negative integers less than $(k+1)$.

This min operator applied to the generating form of the digraphs gives the *generating form of the functional digraphs*, provided we use the Boolean sums. We obtain the function, as usual, by abstracting relative to $z_0 \dots z_{k-1}$ the corresponding form

$$d_k^f = \lambda \vec{z}_{(k)} \left(\prod_{h=0}^{k-1} \left(\sum_{j=0}^{k} z_h^j \right) \right) = \lambda \vec{z}_{(k)} \left(\prod_{h=0}^{k-1} \frac{1 - z_h^{k+1}}{1 - z_h} \right). \tag{2.18}$$

2.4. Integer vectors with restrictions

From the generating form (2.8) of the list of numbers, we immediately obtain the *generating function* v_k *of the vectors with k non-negative integer components*:

$$v_k = \lambda \vec{z}_{(k)} \left(\prod_{h=0}^{k-1} \frac{1}{1 - z_h} \right), \tag{2.19}$$

where the representation of the vector $\leqslant x_0, \dots, x_{k-1} \geqslant$ is simply $z_0^{x_0} \cdot \dots \cdot z_{k-1}^{x_{k-1}}$, since the addition of 1 to the last component is not necessary knowing the length, k, of the vector.

Imposing the condition that the h-th component of the vector be less than y_h is equivalent to truncating the development in series of $\dfrac{1}{1 - z_h}$ to the first y_h terms. Therefore the *generating function* $v_{\vec{y}_{(k)}}$ *of vectors whose h-th component is a non-negative integer less than* y_h is:

$$v_{\vec{y}_{(k)}} = \lambda \vec{z}_{(k)} \left(\prod_{h=0}^{k-1} \left(\sum_{j=0}^{y_h - 1} z_h^j \right) \right) = \lambda \vec{z}_{(k)} \left(\prod_{h=0}^{k-1} \frac{1 - z_h^{y_h}}{1 - z_h} \right). \tag{2.20}$$

In the particular case where $y_0 = y_1 = \dots = y_{k-1} = k+1$, we meet again the generating function $\boldsymbol{d}_k$ of the functional digraphs with k vertices.

If we make the substitution in the form corresponding to the function (2.19):

$$z_h = \prod_{r=h}^{k-1} t_r, \tag{2.21}$$

the generic term $z_0^{x_0} \cdot z_1^{x_1} \cdot \dots \cdot z_{k-1}^{x_{k-1}}$ becomes $t_0^{x_0} \cdot t_1^{x_0 + x_1} \cdot \dots \cdot t_{k-1}^{x_0 + x_1 + \dots x_{k-1}}$. By abstracting relative to the new variables $t_0 \dots t_{k-1}$ we obtain the *generating function* $\overset{\geqslant}{\boldsymbol{v}}_k$ *of vectors with k non-negative and non-decreasing integer components*

$$\overset{\geqslant}{\boldsymbol{v}}_k = \lambda \vec{t}_{(k)} \left(\prod_{h=0}^{k-1} \frac{1}{1 - \prod_{r=h}^{k-1} t_r} \right). \tag{2.22}$$

From a vector with non-decreasing integer components we can obtain a vector with increasing integer components by adding, to the h-th component, $h-1$. This corresponds to multiplying each addendum of the generating form by $\prod_{h=0}^{k-1} t_h^h$. Then the *generating function* $\overset{>}{\boldsymbol{v}}_k$ *of the vectors with non-negative increasing integer components* is

$$\overset{>}{\boldsymbol{v}}_k = \lambda \vec{t}_{(k)} \left(\prod_{h=0}^{k-1} \frac{t_h^h}{1 - \prod_{r=h}^{k-1} t_r} \right). \tag{2.23}$$

2.5. Integer multisets with restrictions

The *generating function* $\boldsymbol{m}_{<k+1}$ *of multisets of non-negative integers less than* $(k+1)$ is obtained from the generating function m by limiting from above the index h:

$$\boldsymbol{m}_{<k+1} = \lambda z \left(\prod_{h=0}^{k} \frac{1}{1 - z^{z^h}} \right). \tag{2.24}$$

Imposing the condition that the multiplicity of the number h in the multiset be less than y_h is equivalent to truncating the development is series of $\frac{1}{1-z^{z^h}}$ at the first y_h addenda. Therefore the *generating function* $\boldsymbol{m}_{\vec{y}_{(k)}}$ *of the multisets of non-negative integers less than* $(k+1)$ *in which the multiplicity of the number h is less than* y_h is

$$\boldsymbol{m}_{\vec{y}_{(k)}} = \lambda z \left(\prod_{h=0}^{k-1} \left(\sum_{j=0}^{y_h - 1} z^{j z^h} \right) \right) = \lambda z \left(\prod_{h=0}^{k-1} \frac{1 - z^{y_h z^h}}{1 - z^{z^h}} \right). \tag{2.25}$$

2.6. A one-one correspondence between vectors and multisets

The well known one-one correspondence — Parikh's mapping — (see [6] p. 146) between multisets whose elements are non-negative integers less than k and vectors with k integer components, indicated by

$$[\underbrace{0, \ldots, 0}_{x_0}, \underbrace{1, \ldots, 1}_{x_1}, \ldots, \underbrace{k-1, \ldots, k-1}_{x_{k-1}}] \leftrightarrow \langle x_0, x_1, \ldots, x_{k-1} \rangle \tag{2.26}$$

leads to the following correspondence between the representation as monomials

$$z^{x_0 + x_1 z + \ldots + x_{k-1} z^{k-1}} \leftrightarrow z_0^{x_0} \cdot \ldots \cdot z_{k-1}^{x_{k-1}}. \tag{2.27}$$

This correspondence is equivalent to the substitution of z^{z^i} by z_i. In fact, with this substitution we can obtain from the generating form $\boldsymbol{m}_{<k}(z)$ the generating form $\boldsymbol{v}_k(\vec{z}_{(k)})$.

3. Listing information structures

Linsting the elements of a family of information structures is based on the following *ordering principle* [3]: "The one-one correspondence between a given set of structures and the non-negative integers can be established by creating a total ordering among the elements of the set, that is, the components of the generating function of the same".

Let each element of the family be uniquely determined by the values of n parameters (non-negative integers) which represent its properties. Let the element whose parameters are described by $\vec{x}_{(n)}$ be represented by the monomial $\prod_{j=0}^{n-1} t_j^{x_j}$.

As a first step in the ordering, let us, in fact, decompose the family according to the number, n, of parameters needed to specify each element. The generating form $\mathscr{H}(\vec{t})$, can then be considered as a sum of forms

$$\mathscr{H}(\vec{t}) \stackrel{\mathrm{df}}{=} \sum_{n=0}^{\infty} \mathscr{H}_n(\vec{t}_{(n)}). \tag{3.1}$$

The problem, now, is to order the elements produced by a single generating function $\mathscr{H}_n$. We may assume that the relative order of the elements reflects the alphabetic order of the vectors formed by the parameters which denote the elements. At this point, it is sufficient to specify a formula that gives for each element of the family under consideration the number of elements less than it. For this purpose we define

$$\mathscr{H}_n(\vec{t}_{(n)}) \stackrel{\mathrm{df}}{=} \sum_{j_{n-1}=0}^{\infty} \cdots \sum_{j_h=0}^{\infty} \mathscr{H}_{h,\vec{j}_{(h,n)}}(\vec{t}_{(h)}) \prod_{r=h}^{n-1} t_r^{j_r}, \qquad 0 \leq h \leq n-1; \tag{3.2}$$

$$\mathscr{H}_n^{\vec{x}(n)}(\vec{t}_{(n)}) \stackrel{\mathrm{df}}{=} \sum_{j_{n-1}=0}^{x_{n-1}-1} \cdots \sum_{j_0=0}^{x_0-1} \mathscr{H}_{0\vec{j}_{(n)}} \prod_{r=0}^{n-1} t_r^{j_r}. \tag{3.3}$$

Note that for $h=0$ in (3.2)

$$\mathscr{H}_n(\vec{t}_{(n)}) = \sum_{j_{n-1}=0}^{\infty} \cdots \sum_{j_0=0}^{\infty} \mathscr{H}_{0,\vec{j}_{(n)}} \prod_{r=0}^{n-1} t_r^{j_r}. \tag{3.4}$$

That is $\mathscr{H}_{0,\vec{j}_{(n)}}$ equals 1 or 0 according to whether or not the element denoted by the vector of the parameters $\vec{j}_{(n)}$ belongs to the set $\mathscr{F}_n$ generated by the function $\mathscr{H}_n$.

The meaning of $\mathscr{H}_n^{\vec{x}(n)}(\vec{t}_{(n)})$ is the generating function of the set of the vectors $\mathscr{F}_n$ alphabetically less than the vector $\vec{x}_{(n)}$. Its cardinality is therefore the very number that corresponds to the element denoted by that vector of parameters. Let us observe that if $\mathscr{H}_n$ is a sum of monomials in $\vec{t}_{(n)}$, its cardinality is obtained by substituting $t_0 = \ldots = t_{n-1} = 1$. In our case we have

$$|\mathscr{H}_n^{\vec{x}(n)}(\vec{t}_{(n)})| = \mathscr{H}_n^{\vec{x}(n)}(1_{(n)}). \tag{3.5}$$

Using the further form $\mathscr{H}_{h,\vec{j}_{(h,n)}}^{x_{h-1}}(\vec{t}_{(h)})$ defined by

$$\mathscr{H}_{h,\vec{j}_{(h,n)}}^{x_{h-1}}(\vec{t}_{(h)}) \stackrel{\mathrm{df}}{=} \sum_{j=0}^{x_{h-1}-1} \mathscr{H}_{h-1,\vec{j}_{(h,n)},j}(\vec{t}_{(h-1)})\, t_{h-1}^{j}, \quad 1 \leq h \leq n, \tag{3.6}$$

we give an expression of $\mathscr{H}_n^{\vec{x}(n)}$ which will be found more convenient than (3.3) in the calculations.

$$\mathscr{H}_n^{\vec{x}(n)}(\vec{t}_{(n)})=\sum_{h=1}^{n}\prod_{r=h}^{n-1} t_r^{x_r}\,\mathscr{H}_{h,\vec{x}(h,n)}^{x_h-1}(\vec{t}_{(h)}). \tag{3.7}$$

Equating the cardinalities of both members we have

$$|\mathscr{H}_n^{\vec{x}(n)}(\vec{t}_{(n)})|=\sum_{h=1}^{n}|\mathscr{H}_{h,\vec{x}(h,n)}^{x_h-1}(\vec{t}_{(h)})|. \tag{3.8}$$

The number ζ which corresponds to the element identified by the vector $\vec{x}_{(n)}$ is therefore

$$\zeta=\sum_{h=1}^{n}|\mathscr{H}_{h,\vec{x}(h,n)}^{x_h-1}(\vec{t}_{(h)})|=\sum_{h=1}^{n}\mathscr{H}_{h,\vec{x}(h,n)}^{x_h-1}(1_{(h)}). \tag{3.9}$$

The inversion formula may be written in a recurrent form:

$$x_{h-1}=\min j\left[\zeta<\sum_{r=h}^{n-1}\mathscr{H}_{r+1,\vec{x}(r+1,n)}^{x_r}(1_{(r+1)})+\mathscr{H}_{h,\vec{x}(h,n)}^{j+1}(1_{(h)})\right],\quad 1\leqslant h\leqslant n. \tag{3.10}$$

Note that, in fact, ζ is defined for any vector, even if it does not belong to the set $\mathscr{F}_n$. For the actual calculation of ζ as a function of the vector $\vec{x}_{(n)}$ we therefore need an explicit expression for the coefficients $\mathscr{H}_{h,\vec{j}(h,n)}(1_{(h)})$. Owing to (3.2) these coefficients can be obtained by evaluating the partial derivatives:

$$\mathscr{H}_{h,\vec{j}(h,n)}(1_{(h)})=\prod_{r=h}^{n-1}\frac{1}{j_r!}\left[\frac{\partial^{j_{n-1}}}{\partial t_{n-1}^{j_{n-1}}}\cdots\frac{\partial^{j_h}}{\partial t_h^{j_h}}\mathscr{H}_n(1_{(h)},\vec{t}_{(h,n)})\right]_{t_h=,,,=t_{n-1}=0}. \tag{3.11}$$

Generally, however, the calculations are considerably simplified when we are able to write a recurrente equation giving $\mathscr{H}_n(\vec{t}_{(n)})$ as a function of $\mathscr{H}_h(\vec{t}_{(h)})$ and $\vec{t}_{(h,n)}$ for $1\leqslant h\leqslant n$:

$$\mathscr{H}_n(\vec{t}_{(n)})=F(\mathscr{H}_h(\vec{t}_{(h)}),\vec{t}_{(h,n)}). \tag{3.12}$$

In fact, by substituting (3.2) in this equation we obtain $\mathscr{H}_{h,\vec{j}(h,n)}(\vec{t}_{(h)})$ as a function of $\mathscr{H}_{h,f(\vec{j}(h,n))}(\vec{t}_{(h)})$ where the appearance of the function f depends on that of F. In this case it is sufficient to calculate

$$\mathscr{H}_{h,k}(1_{(h)})=\frac{1}{k!}\left[\frac{\partial^k}{\partial t_h^k}\mathscr{H}_{h+1}(1_{(h)},t_h)\right]_{t_h=0} \tag{3.13}$$

to determine the integer ζ corresponding to $\vec{x}_{(n)}$ using the equations (3.9) and (3.6).

3.1 Listing integer vectors

In the case of the family of vectors with non-negative increasing integer components, whose generating function is:

$$\vec{v}_k=\lambda\vec{t}_{(k)}\left(\prod_{h=0}^{k-1}\frac{t_h^h}{1-\prod_{r=h}^{k-1}t_r}\right),$$

the following recurrence equation for the generating forms is obviously valid

$$\vec{\overset{>}{v}}_n(\vec{t}_{(n)})=\overset{>}{v}_{h+1}\left(\vec{t}_{(h)}, \prod_{r=h}^{n-1} t_r\right) \prod_{r=h+1}^{n-1} \frac{t_r^{r-h}}{1-\prod_{s=r}^{n-1} t_s}, \; 0\leqslant h\leqslant n-1. \tag{3.14}$$

Similarly to (3.2) we can develop $\overset{>}{v}_n(\vec{t}_{(n)})$ in a power series in $\vec{t}_{(h,n)}$:

$$\overset{>}{v}_n(\vec{t}_{(n)})=\sum_{j_{n-1}=0}^{\infty} \dots \sum_{j_h=0}^{\infty} \overset{>}{v}_{h,\vec{j}_{(h,n)}}(\vec{t}_{(h)}) \prod_{r=h}^{n-1} t_r^{j_r}. \tag{3.15}$$

By substituting the (3.15) in (3.14) we have

$$\sum_{j_{n-1}=0}^{\infty} \dots \sum_{j_h=0}^{\infty} \overset{>}{v}_{h,\vec{j}_{(h,n)}}(\vec{t}_{(h)}) \prod_{r=h}^{n-1} t_r^{j_r} = \sum_{j=0}^{\infty} \overset{>}{v}_{h,j}(\vec{t}_{(h)}) \prod_{r=h}^{n-1} t_r^j t_r^j \prod_{r=h+1}^{n-1} \frac{t_r^{r-h}}{1-\prod_{s=r}^{n-1} t_s},$$

$$0\leqslant h\leqslant n-1$$

from which, by equating the coefficients of $\prod_{r=h}^{n-1} t_r^{j_r}$, we have

$$\overset{>}{v}_{h,\vec{j}_{(h,n)}}(\vec{t}_{(h)})=\overset{>}{v}_{h,\vec{j}_h}(\vec{t}_{(h)}), \; 0\leqslant h\leqslant n-1 \tag{3.16}$$

and therefore we obtain

$$\overset{>}{v}_{h,\vec{j}_{(h,n)}}(1_{(h)})=\overset{>}{v}_{h,j_h}(1_{(h)}), \; 0\leqslant h\leqslant n-1. \tag{3.17}$$

We then calculate

$$\overset{>}{v}_{h,k}(1_{(h)})=\frac{1}{k!}\left[\frac{\partial^k}{\partial t_h^k} \overset{>}{v}_{h+1}(1_{(h)}, t_h)\right]_{t_h=0}=\frac{1}{k!}\left[\frac{\partial^k}{\partial t_h^k} t_h^h \frac{t_h^h}{(1-t_h)^{h+1}}\right]_{t_h=0}$$

by the Leibniz formula

$$\overset{>}{v}_{h,k}(1_{(h)})=\frac{1}{k!}\sum_{j=0}^{k}\binom{k}{j}\left[\frac{\partial^j}{\partial t_h^j} t_h^h \frac{\partial^{k-j}}{\partial t_h^{k-j}} \frac{1}{(1-t_h)^{h+1}}\right]_{t_h=0}=\binom{k}{h}, \tag{3.18}$$

since the only term which is not eliminated is that in which $j=h$. By analogy with (3.6) we define

$$\overset{>}{v}{}^{x_{h-1}}_{h,\vec{j}_{(h,n)}}(1_{(h)}) \overset{\text{df}}{=} \sum_{j=0}^{x_{h-1}-1} \overset{>}{v}_{h-1,\vec{j}_{(h,n)},j}(\vec{t}_{(h-1)})\, t_{h-1}^j \quad 1\leqslant h\leqslant n. \tag{3.19}$$

By using (3.17) and (3.18)

$$\overset{>}{v}{}^{x_{h-1}}_{h,\vec{j}_{(h,n)}}(1_{(h)})=\sum_{j=0}^{x_{h-1}-1} v_{h-1,j}(1_{(h)})=\binom{x_{h-1}}{h}, \tag{3.20}$$

from which we obtain the following *listing formula for the family of vectors with increasing non-negative integer components*:

$$\zeta=\sum_{h=1}^{n} \overset{>}{v}{}^{x_{h-1}}_{h,\vec{j}_{(h,n)}}(1_{(h)})=\sum_{h=1}^{n}\binom{x_{h-1}}{h}. \tag{3.21}$$

For vectors with non-decreasing integer components, from the generating function (2.22), we deduce the recurrence equation for the generating form:

$$\overset{\geqslant}{v}_n(\vec{t}_{(n)})=\overset{\geqslant}{v}_{h+1}\left(\vec{t}_{(h)},\prod_{r=h}^{n-1} t_r\right)\prod_{r=h+1}^{n-1}\frac{1}{1-\prod_{s=r}^{n-1} t_s}, \qquad 0\leqslant h\leqslant n-1. \tag{3.22}$$

Following the procedure used in the preceding example, we obtain the following *listing formula for the family of vectors with non-decreasing non-negative integer components*:

$$\zeta=\sum_{h=1}^{n}\binom{x_{h-1}+h-1}{h}. \tag{3.23}$$

Finally, we calculate the listing formula in the case of vectors of non-negative integers whose h-th component is less than y_h. Their generating function is:

$$v_{\vec{y}_{(n)}}=\lambda\vec{t}_{(n)}\left(\prod_{r=0}^{n-1}\frac{1-t_r^{y_r}}{1-t_r}\right)$$

from which

$$v_{\vec{y}_{(n)}}(\vec{t}_{(n)})=v_{\vec{y}_{(h)}}(\vec{t}_{(h)})\prod_{r=h+1}^{n-1}\frac{1-t_r^{y_r}}{1-t_r}, \quad 0\leqslant h\leqslant n-1. \tag{3.24}$$

From this recurrence relation we can determine one for the coefficients of the series expansion of $v_{\vec{y}_{(n)}}$. In this expansion the index j_r obviously varies between 0 and y_{r-1}:

$$v_{\vec{y}_{(n)}}=\sum_{j_{n-1}=0}^{y_{n-1}-1}\cdots\sum_{j_h=0}^{y_h-1} v_{\vec{y}_{(n)},\vec{j}_{(h,n)}}(\vec{t}_{(h)})\prod_{r=h}^{n-1} t_r^{j_r}. \tag{3.25}$$

Substituting in (3.24) and equating the coefficients of $\prod_{r=h}^{n-1} t_r^{j_r}$, we have

$$v_{\vec{y}_{(h)},\vec{j}_{(h,n)}}(\vec{t}_{(h)})=v_{\vec{y}_{(h)},\vec{j}_h}(\vec{t}_{(h)}). \tag{3.26}$$

In this particular case the explicit expression of $v_{\vec{y}_{(h)},k}(\vec{t}_{(h)})$ can be deduced from the generating form without derivations. We have in fact

$$v_{\vec{y}_{(h+1)}}(\vec{t}_{(h+1)})=\sum_{k=0}^{y_h-1} v_{\vec{y}_{(h)},k}(\vec{t}_{(h)})\, t_h^k=\prod_{r=0}^{h}\frac{1-t_r^{y_r}}{1-t_r}$$

from which

$$v_{\vec{y}_{(h)},k}(\vec{t}_{(h)})=\prod_{r=0}^{h-1}\frac{1-t_r^{y_r}}{1-t_r}$$

and, therefore

$$v_{\vec{y}_{(h)},k}(1_{(h)})=\prod_{r=0}^{h-1} y_r. \tag{3.27}$$

We deduce, then, the following *listing formula for the family of vectors of non--negative integers with the r-th component less than* y_r

$$\zeta = \sum_{h=1}^{n} x_{h-1} \prod_{r=0}^{h-2} y_r. \tag{3.28}$$

3.2. Listing of sets of integers

The expression of the generating function $s_{<b}$ of the family of sets (empty set included) of non-negative integers less than b is obviously

$$s_{<b} = \lambda z \left(\prod_{r=0}^{b-1} (1+z^{z^r}) \right). \tag{3.29}$$

This expression is formally different from the one dealt with at the begin of this section. We can reduce it to the same shape by the substitution $z^{z^h} = t_h$. This substitution, which corresponds to Parikh's mapping, reduces the problem to that of listing the vectors with binary components, whose solution is given in Sec. 3.1.

Below we list the family of sets of non-negative integers, without imposing upper bounds, as an example of a particulary simple case in which it is neither necessary to represent each elements by a list of parameters nor to subdivide the elements according to the number of parameters necessary to identify them.

We can write the generating form of the sets (empty set included) of non-negative integers (2.10) as a series expansion in which the coefficient of z^{z^j} depends only on j and z:

$$s(z) = 1 + \sum_{j=0}^{\infty} s_{<j}(z)\, z^{z^j}. \tag{3.30}$$

If we define, by analogy.with (3.6)

$$s^{x_{h-1}}(z) \overset{\text{df}}{=} \sum_{j=0}^{x_{h-1}-1} s_{<j}(z) \tag{3.31}$$

we obtain

$$s^{x_{h-1}}(1) = \sum_{j=0}^{x_{h-1}-1} \prod_{r=0}^{j-1} 2 = 2^{x_{h-1}}. \tag{3.32}$$

Then the *listing formula for the sets of integers* is simply

$$\zeta = \sum_{h=1}^{k} 2^{x_{h-1}}, \tag{3.33}$$

which maps $\{x_0, \ldots, x_{k-1}\}$ into ζ.

The correspondence (2.27) may be also used to reduce the listing of multisets of integers to that of the vectors.

4. Conclusion

It might seem that the listing formula (3.33) signifies only the system of binary numeration. In fact, a numeration system does underlie each one-one correspondence with the integers given in this paper. However, the most novel and interesting interpretation of these one-one correspondences is that of using them for the codification of certain data structures, including their restrictions. Using the methods of this paper, based on the generating function, this coding becomes automatic. The study of the generating function has proved, then, quite useful in dealing with data structures. A further development might be an investigation into the meaning of the substitution of forms for variables in the generating functions.

References

1. Böhm C., The CUCH as a formal and description language, In: Formal languages description languages for computer programming. Ed. T.B. Steel. Amsterdam 1966.
2. Böhm C., Structure informative e loro transformazioni. Atti del Seminario sul trattamento dell'Informazione. Instituto U. Dini, Firenze 1971.
3. Böhn C., Ronchi della Rocca S., Numbering methods of finite sets and multisets through trees. Comm. of ACM Symp., Venice 1972.
4. Chomsky N., Schützenberger M. P., The algebraic theory of context-free languages. In: Computer programming and formal systems. Eds. H. Braffort and H. Hirschberg. Amsterdam 1963.
5. Knuth D. E., The art of computer programming. 1. Fundamental algorithms. 2. Seminumerical algorithms. New York 1968—1969.
6. Ginsburg S., The mathematical theory of context free languages. New York 1966.
7. MacMahon M. P. A., Combinatory analysis. New York 1960.
8. Perkus J. K., Combinatorial methods. New York 1971.
9. Riordan J., An introduction to combinatorial analysis. New York 1967.

Struktury danych i ich przekształcenia

Celem artykułu jest zdefiniowanie wystarczająco ogólnej struktury informacji, która umożliwiłaby otrzymanie za pomocą kolejnych klasyfikacji wielu różnych struktur informacyjnych, jak na przykład listy, zbiory lub zbiory wielokrotne. Strata informacji odpowiadająca każdej wymienionej klasyfikacji jest badana z punktu widzenia zachowanych własności. Przedstawiono pewne operacje, które nie zmieniają tych własności. Poza tym przedstawiono powiązania między różnymi rodzinami struktur bez straty informacji, tj. odwzorowania jeden do jednego.

Структуры данных и их преобразования

Целью статьи является определение достаточно общей структуры информации, которая позволила бы получить с помощью очередных классификаций многие разные информационные структуры, такие как списки, множества и многократные множества.

Потеря информации, соответствующая каждой из указанных классификаций, исследуется с точки зрения сохраняемых свойсцв, Представлены некоторые операции ненарушающие эти свойства. С другой стороны представлены связи между разными семействами структур без потери информации, т.е. отображения один к одному.