

Specification of Computer Programs.

C.A.R. Hoare.

E.C. Luscher

~~S.R.C. Senior Fellow~~

Department of Computer Science
The Queen's University, Belfast

The logical
Systems of
Lesniewski

October, 1976

North-Holland 1962

Stapecki S. Lesniewski's
calculus of names. *Studia
Logica*

Summary. A notation is suggested for the rigorous
~~form~~ specification of a problem preparatory to its
solution by computer program. It is also intended
for the specification of subproblems to be solved
by the parts of a large program, and for a
proof that the combination of the parts in fact
solves the original problem. The notations are
selected ~~as~~ to express a subset of the
concepts of mathematics and logic, and a superset
of the facilities of a general-purpose programming language.

1956?

1. Introduction.

A computer program is, in a sense, a complete and rigorous specification of the problem which it is intended to solve. However, it is for most purposes quite unsuitable; it appears far too late in the timescale of a project; it is ~~far~~ too long and detailed; ~~so~~ that it is impossible to check intuitively its correspondence with the intentions of its user; ~~and it appears far too late in the timescale of a project,~~ ~~and it is~~ ~~very~~ difficult to change ^{when} if it is ~~wrong.~~ does not

For these reasons, there is a clear need for some rigorous method of specifying the solution of a programming problem in a manner independent of the program itself. Such a specification should be ^{highly} abstract, ⁱⁿ terms, in order to avoid the ^{voluminous} lengthy detail of computer programs; it should be constructed as the first stage in the development of a program; and it should be capable of intuitive validation by the user of the program, in the same way as an architect's plans can be checked by his ^(discerning) client.

In addition, the initial problem specification ~~should~~ ^{may} be ~~the first step in the~~ ^{suggestive of the subsequent} development of the program, by a systematic progression of more detailed design decisions. At each subsequent stage, the same specification technique should be used to describe the function of each part of the program, and the interfaces between them. Finally, the specification should serve as the formal criterion of correctness of a program and of its parts; so that it is possible to prove, with any desired degree of formality, that a program ^{contains no logical errors,} meets its specification.

A notation for expressing such specifications has been known as a "specification language", an "assertion language" or even a "system description systems analysis language". Many "languages" have been proposed for this role, and are acknowledged in the references. This paper attempts a synthesis of the best ideas of its predecessors, but attempts to improve upon them by avoiding all concepts and notations which more properly belong to the programming language itself.

The design of a "specification language" seems to give even more scope to the whims of its designer than the design of a conventional programming language, since there is no longer the ~~poss~~ obligation of efficient implementation to restrain his ^(more elaborate) flights of fancy. For this reason, the design of a specification of language should be judged most strictly by the remaining design criteria listed below:

(1) Simplicity: the ~~sum~~ length of the description, and the number of basic symbols should be less than (say) ALGOL60

(2) Clarity: the language itself should be completely and unambiguously defined by an ^(intuitively acceptable) ~~consistent~~ axiom set. We shall adopt standard predicate calculus with function symbols and equality; also Peano's axioms for number theory, and a Zermelo-Frankel set theory.

~~(3) Modularity: the language should split into the parts of the language should be logically independent, like the branches of mathematics and logic.~~

(4) General purpose: every feature of the language should be relevant for the description of problems in every major computer application area.

(5) Convenience: the notations should be available on conventional typewriters, and in the standard character set for computers. Subscripts and superscripts must be avoided.

(6) Definitional capability: the language should be "extensible" by the normal definitional methods of logic and mathematics.

(7) Good comment conventions: a specification should be an interleaving of formalism ~~text~~ and informal commentary, like a good mathematical text.

(8) Correspondence with mathematical intuition: the normal methods conventions of mathematical reasoning should be valid.

~~(9) Strictly typed structure: the assignment of a unique type to every variable and expression is a practical aid to understanding and reasoning.~~

(9)
(10) Avoidance of defaults: implicit quantifications, default types, etc., must be avoided; they tend to confuse the unsuspecting reader.

It is hoped that the content of this paper will be of interest to programmers, mathematicians, logicians, programming language designers, and hardware designers; in the following ways:

(1) it will enlarge the imaginative and inventive capabilities of programmers by introducing them to the most useful abstractions of mathematics and logic.

(2) it will encourage mathematicians and logicians to regard computer programming as a natural field of application for their theories and methods.

(3) it will inspire programming language designers to maintain design languages in which programs can be developed naturally from their specifications, and can be proved to ~~consistent~~ meet them.

(4) it will challenge hardware designers to provide sound and efficient implementations for more abstract problem-oriented concepts, thereby reducing the number of steps required for preparation of a program for execution by computer.

~~(5) it ~~is~~ may be possible to construct an automatic proof checker for a proof expressed in a notation similar to the one proposed here.~~

However, there are two dangers that must be avoided: ⑦

(1) The programmer's imagination should not be restricted to the concepts of any fixed "language"; he must be prepared to invent, define, and implement whatever new concepts are required by his problem.

(2) The designer of hardware and ^{programming} soft languages should not be too ambitious in the direct implementation of abstract concepts, for which no uniformly satisfactory implementation is possible.

In this design, I have wholly ignored real numbers and their floating point approximations. I do not know enough about this subject to make a ^{reasonable} design proposal.

1. Comments.

$\langle \text{text} \rangle ::= \langle \text{commentary} \rangle \{ \langle \text{formal tests} \rangle \langle \text{commentary} \rangle \}$

$\langle \text{commentary} \rangle ::= \langle \text{comment line} \rangle \{ \langle \text{comment line} \rangle \}$

$\langle \text{formal tests} \rangle ::= \langle \text{formal line} \rangle \{ \langle \text{formal line} \rangle \}$

$\langle \text{formal line} \rangle ::= \langle \text{a ~~non~~ blank line with space in its first column} \rangle$

$\langle \text{comment line} \rangle ::= \langle \text{a blank line} \rangle |$

$\langle \text{a line with a printed character in its first column} \rangle$

A text consists of commentary interspersed with formal tests. In the sequel, only the formal tests will be treated.

(Apart from its role in ^{explaining and} separating formal tests, the commentary ~~phase~~ has no formal significance, and will be ignored hereafter.)

2. Definitions

$\langle \text{definition} \rangle ::= \langle \text{definition} \rangle \{ \langle \text{definition} \rangle \}$
 $\langle \text{formal text} \rangle ::= \langle \text{simple definition} \rangle | \langle \text{subordinate definition} \rangle$
 $\langle \text{definition} \rangle ::= \langle \text{noun} \rangle = \langle \text{formula} \rangle$
 $\langle \text{simple definition} \rangle ::= \langle \text{noun} \rangle = \langle \text{defining formula} \rangle$
 $\langle \text{defining formula} \rangle ::= \langle \text{formula} \rangle$
 $\langle \text{noun} \rangle ::= \langle \text{identifier} \rangle$

~~$\langle \text{subordinate definition} \rangle ::= \langle \text{no} \rangle \langle \text{simple definition} \rangle$~~

A definition introduces a noun to stand for ^{its defining} formula on the right hand side of the definition. The meaning of the text is unchanged if every occurrence of the noun in subsequent ^{other} formulae is replaced by its defining formula, enclosed in brackets, if necessary.

The defining formula of a simple definition must be closed.

All occurrences of nouns in formulae must be replaceable by the rule given above, and the result of replacement must be a valid formulae. Hereafter, we shall describe only formulae in which these replacements have ^(already) been made; we can therefore ignore the existence of nouns and definitions.

* consequently, no noun can be defined directly or indirectly in terms of itself.

The meaning of the symbols is given in the following table.

3. Brackets.

Brackets are used to enclose formulae. A bracketed formula begins with an open bracket and ends with a matching close bracket; the following table gives the matching pairs:

()
[]
{	}
()
[]
{	}

The choice of bracket pairs is arbitrary; by convention, the later pairs are used to enclose larger formulae.

A close bracket must appear in the same line or the same column as its matching open bracket.

All text enclosed in brackets must be written in columns to the right of the opening brackets.

In the syntactic definitions which follow, only round brackets will be used. The round open bracket may be replaced by any of the other open brackets, provided that the corresponding close bracket is replaced by the matching pair.

These principles lead to the following precedence chain. }
 }
 }

* /

+ -

m

w

'

j

n

u

++

--



x

a

P

Q

r

s

F

I

P

R

l

f

s

<

@

e

=

E

≤

<

b

&

v

⇒

≡

,

To indicate precedence between operators of the same precedence, brackets must be used.

4. Sets.

4.1 Syntax

$\langle \text{set definition} \rangle ::= (\langle \text{declarative formula} \rangle \$ \langle \text{proposition} \rangle)$

$\langle \text{declaration} \rangle ::= \langle \text{variables} \rangle : \langle \text{sets} \rangle$

$\langle \text{variable} \rangle ::= \langle \text{identifier} \rangle$

A declarative formula F is one which may contain declarations

$$x_1 : A_1, x_2 : A_2, \dots, x_n : A_n.$$

The variables x_1, x_2, \dots, x_n must be unique.

The scope of these variables is just the following proposition.

Any other occurrence of the same identifiers in F must refer to variables declared in some enclosing formula.

4.2. Semantics. $(F \$ P)$ is the set of all values y for

which there exist $x_1 \in A_1, x_2 \in A_2, \dots, x_n \in A_n$ which satisfy P , and for which $y = F'$, where F' is formed from F by replacing

all $x_i : A_i$ simply by x_i .

?? Abbreviation $x_1 : A_1, x_2 : A_2, \dots, x_n : A_n$ stands for $x_1 : A_1, x_2 : A_2, \dots, x_n : A_n$

Special cases (1) $(x : A \$ P)$ is the subset of all members x of A

such that P

(2) If F contains no declarations, $(F \$ P)$ is empty if P is false (and F is undefined) and F is defined. (and is the singleton set of F if P is true)

4.3. Examples:

$$(x : \mathbb{N} \$ x/3 = 7)$$

is the set consisting of 21, 22, and 23

$$(x : \mathbb{N} + y : \mathbb{N} \$ x = y)$$

is the set of even numbers.

5. Functions.

5.1 Syntax

Same as declarative formula on previous page but with scope to end of function definition?

$\langle \text{function definition} \rangle ::= (\langle \text{definitional formula} \rangle \rightarrow \langle \text{formula} \rangle \S \langle \text{proposition} \rangle)$

5.2 Semantics.

A function definition $(F \rightarrow R \S P)$ is the function

that maps y to z , where there exists an x_1, x_2, \dots, x_n such that P and $y = F'$ and $z = R$.

? $z = R'$

Note that the scope of the declarations in F is both R and P

~~5.3~~

Abbreviation: $(F \rightarrow R)$ stands for $(F \rightarrow R \S \text{true})$

5.3 Examples.

$(x: \mathbb{N} \rightarrow x + y)$ is a function which adds y to to any natural number x .

$(0 \rightarrow 1)$ is the function that maps 0 onto 1 but is undefined elsewhere.

$(x: \mathbb{N}, y: \mathbb{N} \rightarrow x - (x/y) * y)$ is the function x modulo y .

1. Greatest common divisor.

The divisors of a number x are all numbers y which can be multiplied by some number z to give x

$$\text{divisors} = (\{x \in \mathbb{N} \rightarrow \{y \in \mathbb{N} \mid \exists z \in \mathbb{N} \ (y * z = x)\}\})$$

Shouldn't the 'z' be bound more explicitly to the 'y'?

The common divisors of two numbers x and y are in the

intersection of both their sets of divisors; and we require the greatest of them:

$$\text{gcd} = \text{greatest} (\{x \in \mathbb{N}, y \in \mathbb{N} \rightarrow \bigcap (\text{divisors}(x) \cap \text{divisors}(y))\})$$

2. Longest ascending scattered subsequence.

A scattered subsequence of a sequence x will be represented by the set of indices of the selected elements of x . Every such index is less than the length of x :

$$\text{subseq} = \{ (x: \mathbb{Q}^N \rightarrow \mathbb{P} \mid (y: \mathbb{N} \& y < Lx)) \}$$

"subseq(x) is the set of all subsequences of x ."

The ascending subsequences are those whose elements are sorted:

$$\text{ascend} = \{ (x: \mathbb{Q}^N \rightarrow \mathbb{P} \mid (y: \text{subseq}(x) \& A(i: y, j: y) \& E x @ i \leq E x @ j)) \}$$

i.e. if $z \in \text{ascend}(x)$ then for all i and j in z , the i th element of x is not greater than the j th.

We are interested in the length of the longest of them

$$\text{long} = \{ (x: \mathbb{Q}^N \rightarrow \mathbb{P} \mid (L y: (\text{ascend}(x) \& \text{true})) \}$$

3. Equivalence classes.

A relation \sim which is transitive, reflexive, and symmetric is said to be an equivalence relation

$$x: s \sim s \text{ } \& \text{ } x; x \subseteq x$$

$$\text{transitive} \implies (x: s \sim s \text{ } \& \text{ } A(x: s, y: s, z: s \text{ } \& \text{ } y \sim x \text{ } \& \text{ } z \sim y) \implies z \sim x)$$

$$\text{symmetric} \implies (x \text{ } \& \text{ } A(y, z \text{ } \& \text{ } y \sim x \text{ } \& \text{ } z \sim y) \implies z \sim x)$$

$$\text{reflexive} \implies (x \text{ } \& \text{ } A(y: s \text{ } \& \text{ } y \sim x \text{ } \& \text{ } y)) \quad x: s \sim s \text{ } \& \text{ } \frac{\sum s \subseteq x}{\sum s \sim s}$$

an equivalence \implies transitive \wedge reflexive \wedge symmetric

If \sim is an equivalence relation defined on elements of a set S , the equivalence classes of y (wrt \sim) is

equivalences $\implies (x: \text{equivalence} \rightarrow$

$$\{z: P \text{ } \& \text{ } A(u: s, v: s \text{ } \& \text{ } u \sim v) \implies E(w: z \text{ } \& \text{ } u \sim w \text{ } \& \text{ } w \sim v)$$

4. Maximal strong components.

A graph can be represented as a relation between nodes

The transitive closure of this graph is just the ancestral of this relation:

among the nodes of a graph g :

$$\text{equiv} \leftarrow (g: \text{graph} \rightarrow \text{equiv} \cap (g \text{ p-1}))$$

The strong components of a graph are the equivalence classes induced by this relation

$$\text{maximal strong comp.} \leftarrow (g: \text{graph} \rightarrow \text{equiv classes (equiv (g))})$$

Maximal strong components.

A graph is identified as a relationship between its nodes.

The transitive closure of a relation x is the relation which y bears to z if and only if there is a chain

y_1, y_2, \dots, y_n such that $y = y_1$ and $z = y_n$ and

$y_i \underline{x} y_{i+1}$ for $i = 1..n-1$. (note: if $n=1$, $y=z$)

This can be defined more economically, using relational powers:

transitive closure ==

$$(x: s \underline{r} s \rightarrow \underline{\cup} (x \underline{r}^n: \mathbb{N} \text{ true}))$$

Two nodes are strongly connected in a graph if they are related both by the transitive closure and by its converse

strong connection == $(x: s \underline{r} s \rightarrow$

transitive closure(x) \underline{r} transitive closure(x \underline{r}^{-1})

)

Strong connection is an equivalence relation
The strong components of a graph are simply the equivalence classes induced by it.

$$\text{strong components} == (x: s \underline{r} s \rightarrow \text{equivclasses}(\text{strongconnection}(x)))$$

SYMBOL	NAME	READING	TYPES	
			x	y
<u>(1) Natural Numbers.</u>				
$x \cdot y$	product	x times y	NN	NN
x / y	quotient	x over y	"	"
$x + y$	sum	x plus y	"	"
$x - y$	difference	x minus y	"	"
$\min x, y$	minimum	x min y	"	"
$\max x, y$	maximum	x max y	"	"

(2) Structures

x, y	ordered pair	x paired with y	s	t	set
$\perp x$	first alternative	tag one x	s		set
$\perp x$	second alternative	tag two x	t		set
\emptyset	empty sequence	null			Qs
Qx	unit sequence	queue x	s		Qs.
$x \perp y$	concatenation	x joined to y	Qs	Qs	Qs
Lx	sequence length	length of x	Qs		NN

TYPES

(3) Sets

	x	y	result
\emptyset			\underline{P}_s ($x \in s \iff \emptyset \in x$)
$\{x\}$	s		\underline{P}_s ($y: s \iff y = \{x\}$)
$x \cap y$	\underline{P}_s	\underline{P}_s	\underline{P}_s ($z: s \iff z \in x \& z \in y$)
$x \cup y$	"	"	" ($z: s \iff z \in x \vee z \in y$)
$x \setminus y$	"	"	" ($z: s \iff z \in x \& z \notin y$)
\bar{x}	\underline{P}_s		s
$ x $	\underline{P}_s		$\text{NN } \alpha. \forall f: (P_s \Rightarrow \text{NN}) \rightarrow 0 \rightarrow 0. \forall (y: P_s \Rightarrow \alpha \rightarrow f(y) + 1) \cdot \alpha \cdot z \in y$

(4) Types

NN			$\underline{P}_{\text{NN}}$ ($\alpha \in \mathbb{N} \iff y (y = \alpha + 1)$) $\forall (f: P_{\text{NN}} \rightarrow \underline{S} \cup \dots)$
$x \times y$	\underline{P}_s	\underline{P}_t	$\underline{P}(s \times t)$ ($z, w \in s \times t \iff z \in s \& w \in t$) $(y: \text{NN} + 1) \cdot \alpha \cdot z \in y$
$x \cup y$	"	"	$\underline{P}(s \cup t)$ ($z \in s \cup t \iff z \in s \vee z \in t$)
$\ast x$	\underline{P}_s		$\underline{P} \underline{Q}_s$ ($y \in s \iff \alpha$)
$\underline{P} \alpha$	\underline{P}_s		$\underline{P} \underline{P}_s$ ($y \in s \iff \alpha$)
$f: x \rightarrow y$	\underline{P}_s	\underline{P}_t	$\underline{P}(s \rightarrow t)$
$x \Rightarrow y$	\underline{P}_s	\underline{P}_t	$s \rightarrow y$ $\underline{P}(s \times t) \& (z \in s \implies (A_w, v, u \in s, w, v \in x \& u \in s \implies v = u))$
$x \Leftrightarrow y$	"	"	$s \rightarrow y = (x \Rightarrow y) \cap (y \Rightarrow x)$
$\forall x$	$\underline{P}_s \Rightarrow \underline{P}_s$		$\underline{P} \forall x$ $s \in 0 \cup (\alpha \in \text{P} \text{NN} \cdot s \text{ true})$

$\text{NN} (y: P_s \rightarrow y) \cdot A(z \in z \implies y \Rightarrow \alpha(z) \in y)$

(5) Relations.

$F \subseteq X$	Function from sequence	function of	$Q \subseteq S$	$M \Rightarrow S$	$\forall f: Q \subseteq N \rightarrow (0 \rightarrow 0) \cup$ $\{x: \exists s; y: \exists s \rightarrow$
$I \subseteq X$	identity	identity over x	$P \subseteq S$	$S \subseteq S$	$\{y, z \mid y = z \ \& \ y \in x$ $\} \cup \{z \rightarrow y\} \cup P(x)$
$x \subseteq P \cup y$	relational power	x to the y	$S \subseteq S$	NN	$y \cdot X \rightarrow [0 \rightarrow IV] \cup (n+1 \rightarrow f \circ x)$
$x \subseteq P - y$	converse power	x to the minus y	"	NN	$z, w \mid (z, w) \in x \subseteq P \cup$
$\neq x$	ancestral	star x	"	"	
$x \subseteq y$ or $y \subseteq x$	composition	$\{y$ composed with x $\} \subseteq$ then y	$S \subseteq T$	$T \subseteq U$	$(z, w) \in V \mid z, v \in x \ \& \ v, w \in y$
$x \subseteq y$	update	x else y	$S \subseteq T$	$S \subseteq T$	$x \cup (y - D \subseteq D \cup y)$
$D \subseteq x$	domain	domain of x	$S \subseteq T$	$P \subseteq S$	$(y) \subseteq z \mid y, z \in x$
$R \subseteq x$	range	range of x	"	$P \subseteq T$	$(y) \subseteq z \mid z, y \in x$
$x \subseteq y$	image	x of y	$S \subseteq T$	$P \subseteq S$	
$x \subseteq y$	curried relation		S	$(x) \subseteq u \subseteq v$	$(z, t, w) \subseteq ((x, z), w) \subseteq y$
$x \subseteq y$ or $y \subseteq x$	functional application	x at y or y of x	$S \subseteq T$	S	$T(x, y, z) \subseteq x$
$x \subseteq y$	sequence from function		$NN \subseteq S$	NN	$t \subseteq s \mid L_s = n \ \& \ F(s) = x$

$X^T x @ y$: maplist
 $X^T x . y @ z$: componentwise operation
 $\& sum$: S

the x of y

$y \cdot (y^T \rightarrow (0,0 \rightarrow 90) \rightarrow z) \cdot u(z; q; w) \Rightarrow \int (z; q; x(w))$

Q_t

Q_s

$s \Rightarrow t$

componentwise operation

$Q_s (s \Rightarrow t) Q_t$
 Q_u

$\& prod$: P

Ux

PPE

Pt

$(y \cdot E_{z; w} \cdot z, w \cdot x \cdot y \cdot w)$

Nx

"

"

union
 intersection

Mx

PNN

NN

minimum

Wx

"

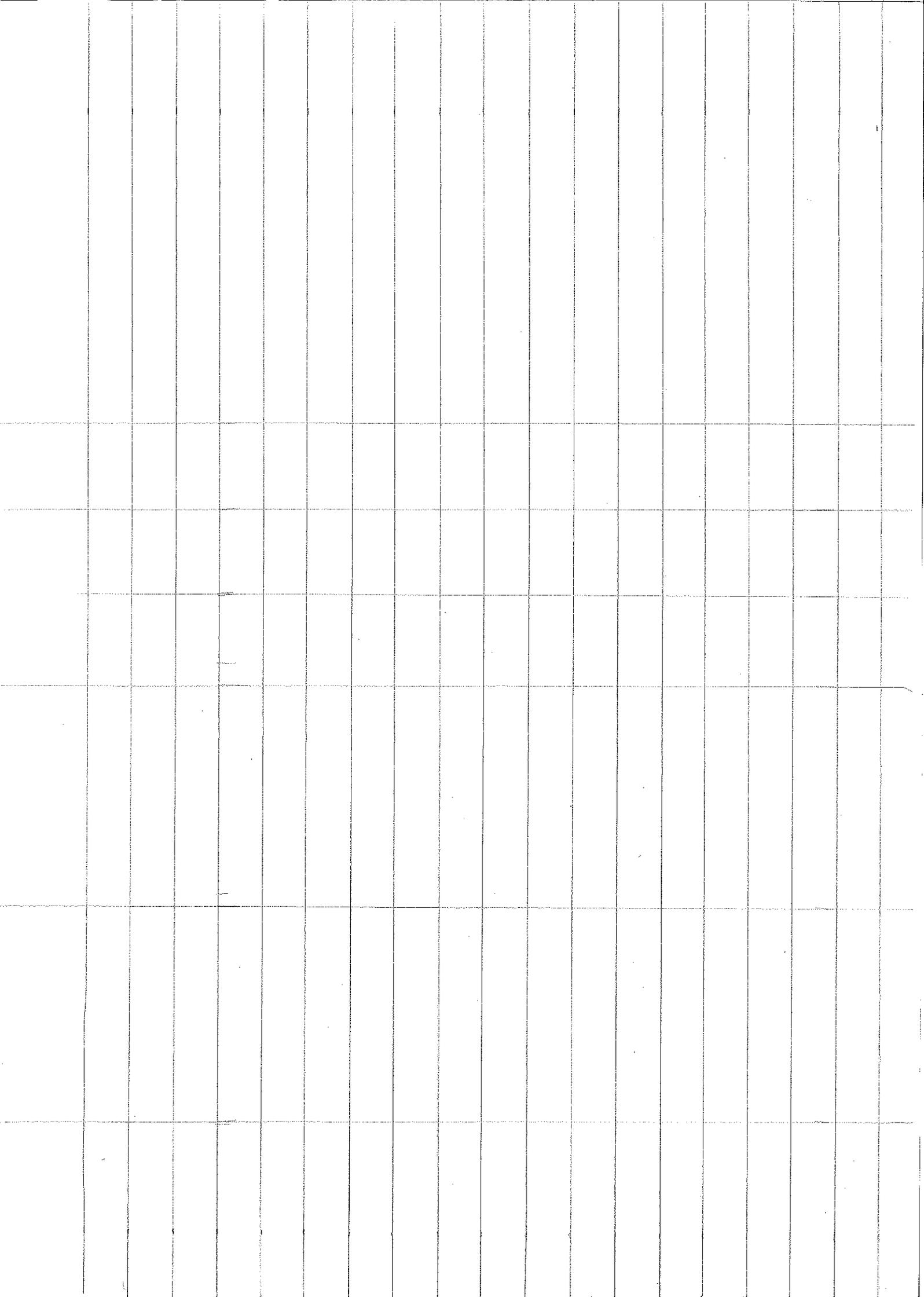
NN

maximum

$x \in y$
 $x = y$
 $x \subseteq y$
 $x \leq y$
 $x < y$
 $x \text{ is a subsequence of } y$
 $x \text{ is a z of } y$

membership	x is a (member of) y	S	P _S	PRED
equation	x equals y	S	S	"
containment	x is a subset of y	P _S	P _S	"
inequality	x does not exceed y	NN	NN	"
strict inequality	x is less than y	"	"	"
initial subsequence	x begins y	Q _S	Q _S	"
infix relation	x is a z of y	S, SRT	S, SRT	"

SYMBOLS	NAME	READING	TYPES		DEFINITION
			op1	op2 result	
$\sim P$	negation	not P	PROP	PROP	$(\sim P) \& (\sim P)$
$P \& Q$	conjunction	P and Q	PROP	PROP	$\sim(\sim P) \& \sim(\sim Q)$
$P \vee Q$	disjunction	P or Q	"	"	$\sim(\sim P \& \sim Q)$
$P \Rightarrow Q$	implication	if P then Q	"	"	$\sim P \vee Q$
$P \Leftrightarrow Q$	equivalence	P if and only if Q	"	"	$(P \Rightarrow Q) \& (Q \Rightarrow P)$
$z \in P$	set abstraction	set of z such that P	S	PROP	P S
$z \rightarrow x$	functional abstraction	function from z to x	S	t	$s \Rightarrow t$
$z \rightarrow x \in P$	restriction	function " such that P	S	t	$s \Rightarrow t$
$A x$	generalisation		\underline{P} S	PROP	$(y \in \sim y \in x) = s \underline{Q}$
$\exists x$	existence		\underline{P} S	PROP	$\sim x = s \underline{Q}$



Index

<u>a</u>	disjoint union
<u>b</u>	begins
<u>c</u>	is contained in
<u>d</u>	
<u>e</u>	is a member of
<u>f</u>	relational image
<u>g</u>	
<u>h</u>	
<u>i</u>	
<u>j</u>	join of sequences
<u>k</u>	
<u>l</u>	else
<u>m</u>	min
<u>n</u>	intersection
<u>o</u>	relational composition
<u>p</u>	relational power
<u>q</u>	sequence defined from function.
<u>r</u>	relation
<u>s</u>	curried relation
<u>t</u>	
<u>u</u>	union
<u>v</u>	or
<u>w</u>	maximum
<u>x</u>	direct product
<u>y</u>	
<u>z</u>	

<u>A</u>	all
<u>B</u>	
<u>C</u>	cardinality
<u>D</u>	domain
<u>E</u>	exists
<u>F</u>	function defined by sequence
<u>G</u>	
<u>H</u>	
<u>I</u>	identity relation
<u>J</u>	singleton sequence
<u>K</u>	
<u>L</u>	length of sequence
<u>M</u>	minimum
<u>N</u>	intersection
<u>O</u>	empty set
<u>P</u>	powerset
<u>Q</u>	quaseset
<u>R</u>	range
<u>S</u>	singleton set
<u>T</u>	the unique
<u>U</u>	union
<u>V</u>	
<u>W</u>	maximum
<u>X</u>	
<u>Y</u>	recursive
<u>Z</u>	