Istituto di Elaborazione della Informazione, CNR, Pisa, Italy

In this paper a self-extensible programming language, SEL or, rather, a general method for a system building tool, is described. The designing criteria are given, the terminology which will be used is specified, the basic concepts of SEL are presented, some typical SEL instructions are illustrated, efficiency and protection problems are discussed, and an implemented version of SEL is outlined. For illustrative purposes, two self-extension programs are given. A short conclusion closes the paper.

(Received May 1970)

1. Introduction

One of the classic ideas for resolving the problem due to programming language proliferation, is to design a universal language. However, until now, although the new 'universal' languages have contributed to clarify many problems, they have complicated many others. We don't expect these difficulties to be completely resolved by the previously conceived self-extensibility as, e.g. ALGOL 68, SIMULA 67, etc. for the following reasons:

- . These universal languages are high level languages. They require a rather complex compiling system which excludes them from the field of small committees
 - them from the field of small computers.

 2. They have an initially defined syntax which, although extensible, binds the system to remain syntax directed. It is not certain that this is the most natural way of describing programming systems, e.g. an operating system.
 - 3. There is no possibility to introduce new basic concepts into the language (e.g. in case of hardware modification) without a drastic modification of the compiling system.
- 4. There is no possibility to extend the language to make it formally equivalent to other, already existing, languages, e.g. FORTRAN, ALGOL 60, etc.
- 5. While there is, in these languages, a general method of compiling, this method cannot be as efficient as the particular one, made for one language.

We consider the most serious restrictions to be points (3) and (4).

In designing SEL, we have taken account of the above criteria. Therefore, SEL is not a high level language, it doesn't use a syntax recompiling mechanism as a built-in facility. Self-extension is procedural, not declarative, so, SEL is also a metalanguage for its own extension. A 'piece of program' can be not only a building element for a further extension but also an operative tool for doing it. The freedom of choosing symbols for denoting new concepts allows the system to describe conventional programming languages in their own syntactic form. Particular care has been taken on efficiency.

Some remarks on terms used

2.1. The term language is used for denoting a tool for man-machine communication which doesn't imply previous

Downloaded from https://academic.oup.com/comjnl/article/14/3/238/420211 by guest on 16 August 2022

- the possibility of extending a language by means of a program written in the same language. Self-extensibility we shall call such before. The term self-extensibility will be used to indicate from the possibility of introducing new basic concepts in the language in the language, or make were not (by modifying its basic implementation) which we shall call capacity to alter also vertical extensibility in order to distinguish it interpretable other messages which the meaning of a message expressed we denote horizontal extensibility. extensibility grammar definition. 2.2. By extensibili pragmatically
- SEL program, i.e. those of its text, are names (words), in agreement with the preference that human beings show towards a text taken as a sequence of words rather than as a sequence of characters. Normally, a man is in fact able to single out a word without necessarily knowing its meaning.

 2.4. In a text a name has two attributes: (i) the commentance of the commentance o
 - Single out a word without necessarily knowing its meaning.

 2.4. In a text a name has two attributes: (i) the occurrence which is its position in the text and (ii) its meaning which is a reference to some physical objects (values) denoted by it.
 - a reference to some physical objects (values) denoted by it.

 2.5. A text of a program—although it may have a meaning—does nothing: it is just a declaration. For associating an action with a text we have to define what we want the program to do. The definition of a program is a mapping of the text into actions. Before executing it, we have to generate the program according to its definition. The generated program can then be executed or, generally speaking, applied. Finally, a program can be

3. Basic concepts of SEL

SEL is self-extensible vertically but, as its structure also facilitates horizontal extensibility, it is to be considered as a general method for building programming systems rather than building a well defined language. The basic instructions

The Computer Journal

presented in this paper are just some examples, and are by no means suggested as an optimal choice.

3.1. A SEL program is an ordered set of objects. An object is composed of two parts: an external object which is the name of the object, and an internal one which is its value. By convention, a SEL object always has two values: an executable one, i.e. a procedure, and a static one of some type (e.g. integer, real, etc.).

One of the basic ideas of SEL is just this active feature of its 'words', i.e. the fact that each object (its procedure) can be executed. To extend this ability to all names, we introduce the convention that to each object not otherwise defined 'belongs' an error message procedure. The internal part of an object can be redefined by applying (executing) other suitable objects (i.e. their procedure). The result of the application of an object is completely defined by:

- The name of the object
- external object (a) its occurrence by which the object is

actually entered

- (b) the meaning of the object (the reference
 - to the value of the object).
 - The value of the object
- (a) procedure

internal object

- (b) static value.
- definition of a SEL program, i.e. the mapping of words into actions is the simplest one-to-one correspondence between external and internal objects. It follows that any arbitrary It is a consequence of what was said before that the sequence of SEL words is a formally correct SEL program.
- Pragmatic errors must be controlled by the program itself, so any 'error message' is to be considered as a result of an execution, not just a message from a supervisor program.

 3.3. The name of an object is singled out from the input character string without taking into consideration its meaning. There is a fixed convention for the inner syntax of words, which cannot be altered by any self-extension. (Although it could be modified by a horizontal extension.) In our implemented version the SEL system recognises identifiers, integer constants, real constants, string constants (all of which are defined as in ALGOL) as well as an additional class of special identifiers and one of special characters.
 - 3.4. SEL has a basic object set (basic vocabulary). The procedures of the basic objects represent the basic actions of SEL. Moreover, each basic object can generate a new internal object (see Section 3.8) which (or a combination of some new internal objects) can be assigned to an external object. Object generation is the main facility of self-extensibility in SEL. 3.5. An object (i.e. its procedure) on conserve that
 - An object (i.e. its procedure) can operate both on its own value and on the ones belonging to other objects, i.e. its parameter can be either itself or another object. As for the basic vocabulary, for the sake of simplicity, we assume that if a procedure has one parameter other than itself, this parameter is the object whose occurrence is successive to the occurrence by which the procedure was entered. For example, if the symbol SPSB has the procedure which prints out the name of a word, the execution of

SPSB APPLE

basic procedure has more than one parameter, these parameters are the objects whose occurrences are successive to the ones by which the procedure was entered. For example the will cause the print out of the symbol APPLE. Similarly, if a

SSUM A B C

C are defined as possessing integer values (see Sections 4.1 and assigns the sum of the static values of A and B to C if A, B and

Parameter evaluation takes part of the procedures of the basic SEL objects by using a system vector (called the occurrence

vector) which contains the external representation of the SEL program in form of pointers to object names. Names contain (point to) values.

set of objects can be considered as one internal object and can be assigned to another external object. One can do this by applying some basic objects (i.e. executing their procedures). E.g. Macros can be easily defined in SEL: an ordered

<* \$PSB APPLE *>
$$\rightarrow$$
 Y

above example, any execution of the procedure of the object. Y' produces the print out of 'APPLE'. In SEL a defined macrocan be executed immediately like a subroutine (as in the above example) or can be generated as a real macro (see Section 3.8) 3.7. Macros may have formal parameters. One way of introducing this is by using an object which is also a macrocanter way of introducing this is by using references, which are a type of value. A particular object, \$ADR has the procedure which, executed, defines its successor as possessing a reference. Another procedure (that of \$ASR) can assign a consists of sequence of objects whose names follow the '<*' until '*>'. The next object to be executed is that following $(*>', in this case '\rightarrow' whose procedure assigns the macro defined to that external object whose name follows it, i.e. to$ 'Y'. Formally, a symbol which denotes a macro is intended to be substituted statically by the sequence of symbols closed between '<*' and '*>'. When the macro is defined as in the above example. assigns the sequence of objects '\$PSB APPLE' to the external object having the name Y. The procedure of the object denoted by the name '<*' defines a macro, i.e. a compound object

assigns a value to R which is a reference to Q. A particular object denoted by '\$REF' is a formal parameter object. Its internal object, the value of the parameter, is the one to which the successor reference definition the execution of the SEL program

\$PSB \$REF R

will cause the print out of 'Q'.

There are two system references accessible to the user symbol to which belongs the macro under execution. E.g. its after the definition.

after the definition

* \$PSB \$REF \$THS *> → T
one executes T, it produces the print-out of its own symbolatie. T. Similarly \$SUC has the value which is a reference to the symbol whose occurrence is successive to the occurrence of the symbol referred by \$THS, by which the macro was called E.g. after the definition

... T W..., it produces the print-out of W. Two particular words \$NXT and \$PRV can cause to move forewards or backwards respectively the system pointer which indicates the occurrence of the symbol pointed out by \$SUC (altering, in if one calls the macro by the occurrence of T in the sequence such a way also the value of \$SUC). This makes possible a look-up from a macro to the 'defining' level.

tion or generation. If the state is interpretation when, by its while, if the state is generation, a new similar procedure is can switch the state of SEL from interpretation to generation The SEL system is always in one of two states: interpretaoccurrence, a basic object is entered, the procedure is executed generated. Particular objects having the name '<:' and or vice versa respectively. Therefore, executing 3.8

$$<: \$PSB \ APPLE :> \rightarrow X$$

8 Number 3 Volume 14

doesn't cause any print out, but entering an occurrence of X after this definition, will result in the print-out of APPLE. (The same symbol '->' could be used for assigning either a macro definition or a macro generation.) A generated macro can have formal parameters as in the case of macro definition. The main difference between the execution of a macro definition and that of a generated macro is that the latter is an optimised procedure while the former is a sequence of calls of distinct procedures under the supervisor SEL system. Note that a macro generated has no internal definition; the entering of the occurrence of the symbol which denotes it will cause the execution of the procedure independently from the state of

Some typical instructions of SEL

- Any symbol can be defined as possessing a value of some type (either variable or constant, see also Section 4.5). The words \$ADR, \$BOL, \$INT, \$REL, and \$CHR define their successor as possessing reference, Boolean, integer, real and duced by field generation and references. On each type of value there are defined some elementary operations: as arithmetic, Boolean operations, assignments of various types, etc. 4.2. The whole program written in SEL is intended to be included between two conventional words: \$PRP and \$SYS. character values. Vectors and structured values can be intro-
- gPRP initially has no procedure, but the user can generate and assign one to it. So each, execution may begin with a user defined preprocessing. \$SYS ends the program execution and, when a conversational version of SEL is available, returns the control to the user. The procedure of the word \$PGO assigns the Boolean value true to its successor if \$SUC has no meaning (it points out of the program). Obviously, it has a practical meaning only internally to a generated procedure.

 4.3. The procedure of \$SKP causes its successor to skip. If the successive symbol denotes a macro of symbols, then the execution of \$SKP causes the whole macro to skip. The words \$SKT and \$SKF cause their second successor to skip if the value of their first successor is Boolean, respectively true
 - or **false.**
- 4.4. The word \$RET causes a return from the execution either of a macro of words or of a procedure. \$REP causes the repetition of the execution of either a group or a procedure.

 4.5. A word can represent either a variable or a constant. By means of the execution of explicit instructions \$VAR or \$CST one can change the state of a symbol into a variable or into a constant respectively. Executing \$CCB assigns the Boolean value true to its second successor if its first successor is a constant, false otherwise. E.g. '->' has the procedure to assign the last defined group or generated procedure to its successor. After the execution of <* \$PSB 'CONSTANT ERROR' \$SYS *> -> ERC \| EXECUTING ERC IT GIVES AN ERROR MESSAGE AND ENDS THE PROGRAM\|

→ \$REF \$SUC <: \$CCB \$REF \$SUC B \$SKF B ERC

THE SAME '→') IS NOT A CONSTANT, IT ASSIGNS THE LAST GENERATED PROCEDURE OR GROUP TO THIS SUCCESSOR I.E. IT DOES THE SAME AS BEFORE IF THE SUCCESSOR OF THE CALLING SYMBOL (OF SAVE THE CONTROL ON CONSTANTS. B IS A BOOL-EAN VARIABLE

The word \$LAB defines its successor as a label pointing to the present occurrence of the symbol referred to by \$SUC. The execution of a label consists in the transfer of the control to the occurrence denoted by the label. E.g. if we wish to define as label any word having an occurrence followed by ':'; pointing to the successor of the same occurrence, we can define the no new symbol group or procedure can be ssigned to a constant. collowing preprocessor:

\$ADR SYM \DEFINES SYM AS POSSESSING A REFERENCE. AS IT IS NOT DEFINED AS A CONSTANT, IT IS

A VARIABLE

WHICH THE PROCEDURE BELONGS TO AND DEFINES THE SYMBOL REFERRED BY SYM AS A THE MACRO LABEL DENOTING THE PRESENT OCCURRENCE SUCCESSOR F SYM *> \rightarrow LB \ADVANCES THE OF THE SUCCESSIVE SYMBOL <* \$NXT \$LAB \$REF SYM *> LB DENOTED BY

<: \$ASA \$SUC SYM \$NXT \ASSIGNS THE REFERENCE VALUE OF \$SUC TO SYM AND ADVANCES \$SUC\ \$CPS \$REF \$SUC:</p>

WHETHER IT IS A ": IF IT IS SO, IT ASSIGNS THE BOOLEAN VALUE TRUE TO B, FALSE OTHERWISEN SISK F B ID COMM. SSKF B LB \SKIP IF THE VALUE OF B IS FALSE, OTHERWISE DEFINE THE LABEL.\

E. ASSIGN | PROTECT \$SPGO B \$SKT B \$REP :> \rightarrow \$PRP \(CONTROL\) WHETHER THE WHOLE PROGRAM IS SEARCHED, REPEAT THE PROCEDURE OTHERWISE. ASSIGN SPRP AGAINST CHANGING PROCEDURE SCST SPRP THE PROCEDURE TO \$PRP

Efficiency and protection

compiler against posterior extension by switching out of the system all basic SEL objects. Using the self-extensible facility of SEL, the volume and complexity of a SEL program (the text of the program) doesn't increase linearly with that of the 5.1. SEL is not designed so that its basic vocabulary can be more elementary than those of a computer, this could result in being too cumbersome. However, an extension of SEL can be tailored to offer a suitable programming language to The interpretation of a program using only the basic instructions of SEL is far from being efficient. However, by The basic SEL instructions are chosen to facilitate this built-in programmed directly. Since its basic instructions are sometimes problem to be described, but less, presumably logarithmically. generating the same program, an optimised routine results. the user. One can also 'protect' the actual extension, e.g. optimisation.

We feel that SEL can produce a program almost as efficient as a program written in machine language for the following reasons:

cedures are easily transformed into a machine language without Basic SEL is an elementary language. Its basic a complicated compile system.

From the practical point of view one of the main sion. The expression view permits an elaborate built in optimisation using fast registers, while in the case of the procedural view it is difficult to distinguish between temporary reasons for considering elements of a language as expressions, instead of elementary operations to be executed, is the problem The solution offered by SEL is the tive. Its value can be referred to only once. E.g. having declared of auxiliary storage necessary for the evaluation of an expresmacro generation, an auxiliary variable is intended as destrucpossibility to declare objects as temporary variables. as auxiliary variable, by executing storage and variables.

the internal object generated by

SSUM A B X SDIV X C D

can be easily optimised by a built in mechanism resulting the object code

DCBA LOAD SUM DIV STO

Due to the low level of basic SEL, it can be easily translated by exploiting the existing machine instructions, e.g. by translating a skip and a jump as a conditional jump; a compare

operation and a skip as a compare-skip instruction, etc. 5.2.3. The possibility of defining an object as a constant has a fundamental role in optimising macro generation. An operation on a constant object is executed during procedure generation. E.g. the object identified by

SREF Y

if Y is a constant, is identified during generation. Otherwise

necessary for the execution of a program. Obviously, the space requirement is not minimal if the program 'uses' only the is identified during execution.

2.4. An external SEL object is composed of a procedure of a store for its static value. We consider the a unique vector field, yielding in such a way, an optimum store required for its name (character string) as not strictly static value of an object. It is true for the basic SEL system. An extension of it can allocate all static values (also types) allocation. entry and 2.4.

a few types of error. Any extension implies new conventions intro-duced into the system, in other words, possibility of new types of error. The strategy of SEL is to offer the means for program error checking. The various levels of boot-strapping allow one to check for any error at the lowest level in which it is detectable; in other words, at compile time or when compiling the compiler, etc. On the other hand, the program itself, which defines self-extension of SEL, shows clearly the kind of possible The basic SEL instructions are protected against errors and how they are checked.

6. Implementation

A first version of SEL has been implemented on the Hewlett Packard 2116/B computer of 8k configuration and has been running since February 1970. Implementation has required 3,000 machine instructions and six man-months. The present version is not highly optimised, because of its experimental character, and, as we had no backing store, because we wanted to reserve more core memory for programs. Based on our experience we believe that a more sophisticated implementation of SEL would require about 4,500 to 5,000 machine instructions (computers with more sophisticated instructions require more sophisticated optimisation; so, the number of instructions required should be more or less the same) and about 6 to 12 man-months. about

The present version is conversational. Some instructions such as delete a line, etc., are included in the basic vocabulary of SEL and therefore can also be generated.

include also instructions for file management. In such a way, In a future more sophisticated version, we should like to operating systems could also be described in SEL.

7. An example of a SEL program

The program in Appendix A defines a self-extension of SEL. After its execution one can declare nouns, adjectives, and other words as pairs of the English and Italian representation. E.g. E.g.

NOUN CAT GATTO, DOG CANE; ADJECTIVE BLACK NERO, BIG GRANDE; WORDS IS E, THE IL;

duces its print-out in Italian and vice versa, taking into account any program composed of declared English words prothat in English the attribute precedes the subject, while in Italian it follows. If the program were

THE BIG DOG IS BLACK

the print-out would be

IL CANE GRANDE E NERO

For simplicity sake, parameters will be indicated as p1, p2 and p3 and they are meant as the first, second and third object successive one referred to

Assigns the static integer value of p_1 to p_2	Compares the name of $p1$ with that of $p2$,	assigns the Boolean result to p3	Compares the procedure of $p1$ with that of $p2$,	assigns the Boolean result to $p3$	Assigns the Boolean value (int. stat. of $p1 =$	int. stat. of $p2$) to $p3$
7	ю		ю		3	
PASI	&CPS		$$^{\text{CPP}}$		SIEQ	

of Assigns the Boolean value (int. stat. int. stat. of p2) to p33 3

SIGT

SILT

units of type p2 Assigns the Boolean value (int. stat. of p1 <Generates a vector field of p3 int. stat. of p2) to p33

\$VEC

and makes p_1 point to this field

Increases the static reference value of p_1 by p_3 units of type p_2 Decreases the static reference value of p_1 by p_2 Assigns the static integer value of p_1 by p_2 3

3 2

SDEC

SINC

SNEG

\$RDI 1 Reads an integer value and assigns it to \$p1\$
\$PVI 1 Prints the static integer value of \$p1\$
\$PTS — Prints a blank
\$PTI — Prints a carriage return and linefeed
\$TXT — Executes the program in the text buffer
\$DTX — Executes the program in the text buffer
\$DTX — Executes the contents of the text buffer
\$Fig. 1. List of some basic SEL objects used in the examples

Or, if the program were

IL GATTO NERO E GRANDE

THE BLACK CAT IS BIG

Note: (a) Any text printed by the user is just a program note; data.

(b) This example of SEL program is not intended topsolve any problem of linguistics.

The program in Appendix B defines and applies an interpretered for a simple subset of ALGOL 60. The interpreter is designed for a 'small' computer.

To help with understanding these programs Fig. 1 gives a list of some of the basic SEL objects used in the examples.

8. Conclusion
Since the basic vocabulary of SEL is not rigorously defined, respectively. SEL is a general method of system building rather than a SEL programming language in the conventional sense. However, it approximates the design criteria in Section 1. As for point (3) in the Introduction, we remark that SEL can be extended also horizontally simply by adding a new object to the basic set. However, the degree of altering the system in this way is not independent of the optimisation technique used.

A criticism of SEL is that, because of its self-extensible

A criticism of SEL is that, because of its self-extensible facility, it requires rather a specialist as a user. In the present This is the reason why we consider SEL as essentially a tool stage, an erroneous program can also break down the system. for system designers.

Acknowledgements

Thanks are gladly due to Prof. A. Caracciolo, to Mr. A. Andronico for stimulating discussions, to Mr. Sprugnoli for supervising and Mrs. Carobbi and Mr. Ricciardi for executing the implementation.

Appendix B