**PCT** WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

**(51) International Patent Classification 4 :** G06F 15/20

A2

**(11) International Publication Number:** WO 86/ 05294

**(43) International Publication Date:** 12 September 1986 (12.09.86)

**(21) International Application Number:** PCT/GB86/00104

**(22) International Filing Date:** 27 February 1986 (27.02.86)

**(31) Priority Application Number:** 8505340

**(32) Priority Date:** 1 March 1985 (01.03.85)

**(33) Priority Country:** GB

**(71) Applicant** *(for all designated States except US):* VICKERS SHIPBUILDING AND ENGINEERING LIMITED [GB/GB]; Barrow Engineering Works, P.O. Box 12, Barrow-in-Furness, Cumbria LA14 1AF (GB).

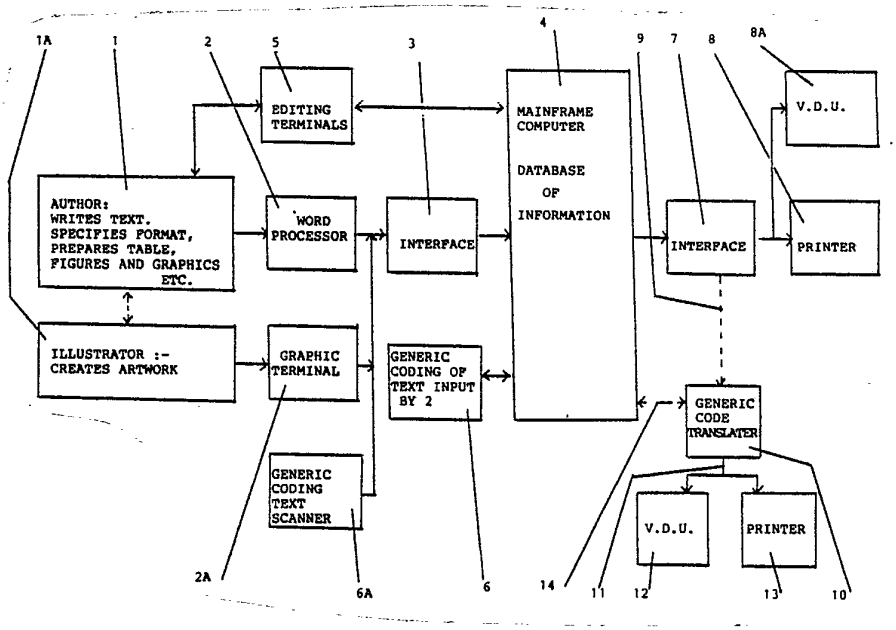
**(72) Inventor; and**
**(75) Inventor/Applicant** *(for US only) :* BOWE, Peter, Karl [GB/GB]; Edenhurst, Eden Mount, Grange-over-Sands, Cumbria (GB).

**(74) Agent:** HASELTINE LAKE & CO.; Hazlitt House, 28 Southampton Buildings, Chancery Lane, London WC2A 1AT (GB).

**(81) Designated States:** AU, DE (European patent), DK, FI, FR (European patent), GB, GB (European patent), IT (European patent), JP, NO, SE (European patent), US.

**Published**
*Without international search report and to be republished upon receipt of that report.*

**(54) Title:** ELECTRONIC TEXT HANDLING




**(57) Abstract**

An electronic text handling apparatus, wherein a source document is constituted as a bare text (24), consisting of the subject-matter of the document free from text formatting instructions, and mark-up codes (23), specifying the text format to be applied to the subject-matter, distinguished from the bare text (24). The apparatus includes processing means (2, 4), having a store holding a mark-up code dictionary defining interpretations of the mark-up codes for specifying text format, able to process the bare text in accordance with the mark-up codes. The source document can be accessed (25) with the bare text formatted according to the interpreted mark-up codes.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|---|---|---|---|---|---|
| **AT** | Austria | **GA** | Gabon | **MR** | Mauritania |
| **AU** | Australia | **GB** | United Kingdom | **MW** | Malawi |
| **BB** | Barbados | **HU** | Hungary | **NL** | Netherlands |
| **BE** | Belgium | **IT** | Italy | **NO** | Norway |
| **BG** | Bulgaria | **JP** | Japan | **RO** | Romania |
| **BR** | Brazil | **KP** | Democratic People's Republic of Korea | **SD** | Sudan |
| **CF** | Central African Republic | | | **SE** | Sweden |
| **CG** | Congo | **KR** | Republic of Korea | **SN** | Senegal |
| **CH** | Switzerland | **LI** | Liechtenstein | **SU** | Soviet Union |
| **CM** | Cameroon | **LK** | Sri Lanka | **TD** | Chad |
| **DE** | Germany, Federal Republic of | **LU** | Luxembourg | **TG** | Togo |
| **DK** | Denmark | **MC** | Monaco | **US** | United States of America |
| **FI** | Finland | **MG** | Madagascar | | |
| **FR** | France | **ML** | Mali | | |

Electronic text handling

The present invention relates to electronic text handling.

A complete text, or source document, prepared electronically, for example on a word processor, is embodied as a series of codes. These codes can be stored, for example on a floppy disk. The codes making up a source document fall into two categories. Codes in one category designate the characters which appear in the text, e.g. letters of the Roman alphabet, Arabic numerals, punctuation marks, spaces etc. Codes in the other category designate the format in which the text is to be presented. That is, these codes provide text output definition: they give instructions as to how the text, or parts of the text, should be presented when the text is displayed or printed-out. These codes may designate, for example, "bold", "underline", "capitals", "centralise", "superscript", "indent", "scientific notation", etc.

Codes of the second category are embedded in the complete electronic text or source document, inserted automatically upon operation of appropriate keys of the word processor when the source document is originally prepared, and invisible when the source document is displayed or printed out.

A major problem in the field of electronic text handling is incompatibility of different text handling systems, such that a source document prepared on one system cannot be transferred to another system.

There are internationally recognised standard codes, such as ASCII and EBCDIC, for characters which routinely appear in a text (e.g. letters of the Roman alphabet, Arabic numerals etc., as mentioned above.

However, even if two systems operate on the basis of the same standard character codes, it is

still highly unlikely that a source document prepared on one system can be transferred to the other without great difficulty.

This is because codes of the second category, output definition instructions, embedded or hidden in a source document, are in general unique to individual systems. Thus, a source document can be readily transferred from one system to another only if the systems are identical or belong both to a group of compatible systems (produced by the same manufacturer).

For transfer of a source document between incompatible systems, even systems using the same standard character codes, it has generally been necessary to obtain hard copy from the first system and for an operator to re-enter the complete text into the second system, by copy-typing from the hard copy.

There have been attempts to produce hardware and/or software to enable source documents to be transferred between different, mutually incompatible, text handling systems, but the results have been expensive, slow and specific to individual pairs of systems.

Put simply, electronic output definition instructions embedded in an electronically prepared text or source document can usually be read only by the system which originally generated the text or by another system specially adapated or programmed. The result is that texts created at one work station system may be inaccessible at another. As the number of different manufacturers offering competing systems has grown, so has this problem of incompatibility.

Another problem with text handling systems arises when it is desired to change the format or output definition of a text. A requirement for a

change of format, whilst retaining the same basic subject matter of a text, occurs frequently in practice; for example the same documentation has to be produced for, say, more than one customer.

In practice, the only presently available means for changing the format of a text is for an operator to go through the text line my line making the necessary alterations. In reality, it may be quicker for the operator to retype the entire text.

Electronic text handling apparatus embodying the present invention can provide for the preparation and handling of an electronic text in such a manner that the problem of incompatibility mentioned above is removed or mitigated.

Electronic text handling apparatus embodying the present invention can provide for the preparation and handling of an electronic text, in such a manner that the problem of format alteration mentioned above is removed or mitigated.

Electronic text handling apparatus embodying the present invention provides that formatting or output definition instructions, hereinafter referred to as mark-up codes or generic codes, are separately and readily distinguishable from the character codes which specify the subject matter, or "bare text", of an electronic text or source document.

That is, a bare text, free from formatting or output definition instructions, and mark-up codes, specifying format or output definition instructions defining a structure for the complete text, can be readily accessed, prepared, handled, modified or edited, essentially separately from one another.

For example, the apparatus allows a bare text to be entered separately from the mark-up codes defining the structure of the complete text.

The mark-up codes may be applied to the bare text as necessary, for example on a line by line basis or at the beginnings of paragraphs etc.

In a complete text, the mark up codes are distinguished from the codes of the bare text by delimiters. For example, the delimiters are special codes for ":" (indicating the start of a mark-up code) and "." (indicating the end of a mark-up code).

Apparatus embodying the present invention can be provided with means for displaying a complete text in two versions.

In the first version, the complete text is displayed to an operator with bare text and mark-up codes separate and distinct from one another.

In the second version, the text is displayed in the format specified by the mark-up codes, i.e. with the proper output definition (with the mark-up codes invisible).

By referring to the first version, the operator can readily access and amend the mark-up codes to effect a change of format which will be visible on the second version.

Apparatus embodying the present invention can provide for easy translation of the separated and distinct mark-up codes to those appropriate for different apparatus.

Apparatus embodying the present invention can provide that the same mark-up codes can be interpreted in different ways.

In effect, the apparatus provides a dictionary by which mark-up codes are interpreted and given meaning (i.e. given meaning in terms of the text format or output definition which they are to specify). The dictionary may include different "profiles". With one "profile", codes registered in the dictionary are

interpreted to have certain meanings, whilst with another "profile" the codes are interpreted to have different meanings.

Thus, by switching "profiles", a text can be provided in different formats without alteration to the mark-up codes related to the text.

Reference is made, by way of example, to the accompanying drawings, in which:-

Figure 1 is a schematic block diagram illustrating electronic text handling in accordance with an embodiment of the present invention, and

Figure 2 shows schematically a dual display of a portion of a text, as provided by an embodiment of the present invention.

Referring to Figure 1, which illustrates the hardware of electronic text handling apparatus according to an embodiment of the present invention, together with the flow of text information associated with the apparatus, normally an author 1 will initially compose a text which is then entered into a word processor 2 by an operator.

The text may be printed off the word processor for the author to correct and amend as necessary for subsequent correction by the word processor operator. In other cases, as shown in the Figure, the text may be passed via an interface 3 to a computer 4 and subsequently accessed by the author e.g. for editing, formatting and printing personally, using an editing terminal 5 for example.

In some cases an illustrator 1A may create artwork etc. to supplement the text. The artwork is prepared on a graphic terminal 2A and submitted to computer 4 via interface 3.

The computer 4 may be a "mainframe computer", but it will be understood that only limited computing

power is required so that a stand-alone word processor could accomplish all the functions so far described. Alternatively a separate microprocessor, personal computer, or other suitable means up to and including a mainframe computer could be used. For large complicated documents, a mainframe computer would be necessary to provide a complete database of information.

At some point in the preparation stage, mark-up codes must be added to bare text. This is normally done by the author 1 or an experienced word processor operator 2 when the text is first created. alternatively, the text could be submitted to the computer 4 which could be used to examine the document and add the appropriate mark-up codes. This second means of adding the mark-up coding is indicated by item 6 on Figure 1. An electronic text comprising bare text and mark-up codes may be derived from a printed document using a text scanner device, similar to a photocopier, 6A which scans, reads the text content and format of the printed document and produces an electronic output. Alternatively, documents already available in electronic form can be coded via item 6.

To explain the nature and use of mark-up codes in an embodiment of the present invention reference will now be made to figure 2, which represents an A3 size Visual Display Unit (VDU) screen. A4 is a commonly used standard size page for reports, etc. The A3 size is double the area of A4 and is equal to two A4 pages placed side-by-side. Thus on the A3 screen 20, an imaginary line 21 divides the screen into two A4 size halves. The advantage of using an A3 size screen is that the mark-up codes and bare text can be displayed on the left-hand side of the screen and the corresponding formatted text, reproduced to the required format according to the mark-up codes, can be displayed

on the right-hand side of the screen, or vice versa, in a dual display. This dual display arrangement is highly advantageous, particularly when used for editing purposes.

On the left-hand side of screen 20 (or on a separate A4 size VDU), the mark-up codes and the bare text are displayed. As an example, this display comprises line numbers 22, mark-up codes 23 and bare text 24. After a line number 22, specific delimiter symbols indicate to the start and end of a mark-up code; in the example shown in Figure 2, a colon (:) indicates the start of a mark-up code stop (.) its end. An actual mark-up code between these delimiter symbols is, for example, a series of letters and/or numbers, the meaning of which is entered as part of a code 'dictionary' in the computer's or processor's memory. In the illustrated example, :T1. (line 30) indicates a title in capitals, centralised on the page and underlined, while :T2. (line 10) indicates the same as T1, but without underlining. Of the other codes, :H4. indicates a heading at the left-hand side of the page in normal text, i.e. only the initial letter would be a capital, and :P5. a paragraph indented 5 spaces. Headings, paragraphs, etc., may be numbered automatically, if required. The text 24 follows after the symbol indicating the end of the mark-up coding; it could follow immediately, as shown, or be separated by a space(s) as required. No mark-up code is required on line 70 as the text would be printed normally until the next mark-up code indicated, for example, the start of the next paragraph.

The mark-up codes are not restricted merely to the placing of titles, etc. on the page. They can be used for a range of other functions such as compiling the document by collecting data from other

sources. For example, on line 140, the code :abcd. could instruct the computer to include Table 1, currently stored on File 8, in the text at that point. Mark-up codes could also be used to add standard items of text to a document, e.g. Copyright infringement warnings, etc.

Where an A3 size VDU screen is used, the completed text 25, formatted according to the instructions dictated by the mark-up codes, is displayed on the right-hand side of the screen. Though the use of an A3 size screen is not essential at the initial word processing input stage 2, it is highly advantageous at the editing stage, where the author can check both for textual and formatting errors simultaneously and correct them on the left-hand side of the screen.

If a document which has not been marked-up with mark-up codes is available, the computer 4 may 'read' the document via interface 3 and add the appropriate mark-up coding. In this case, the text 25 on the right of Figure 2 will be converted to the form 22, 23, 24, as shown on the left.

Once a document has been coded and checked, it may be stored and accessed by computer 4 as and when required. For example, it could be passed via an interface 7 to a printer 8 or a VDU 8A.

If the document is to be used on a second system which uses a different mark-up code dictionary from that with which the document has been prepared, it will be necessary to change either the mark-up codes to those required by the dictionary of the second system or change the dictionary of the second system. As changing the dictionary of the second system may be a lengthy and exacting task, the easier way is to provide translation of mark-up codes from the first type to the second type. This operation is illustrated

diagrammatically by code translator 10 in Figure 1.

If a document prepared in accordance with a first mark-up code dictionary is to be transferred from a first location to a remote second location where the equipment uses a second mark-up code dictionary, the procedure to be used would be as follows:-

A mark-up code translator converts codes of the first dictionary to those of the second. This translation could be effected before transmission to the remote location or after reception at the remote location.

If the translation was to be done before transmission, the document would be processed by translator 10 via connection 14 and then transmitted via an interface (not shown) and an electronic connection 11, e.g. a telephone link, to the remote second location. Thus the transmission path would be from computer 4, via connection 14, translator 10, interface (not shown), electronic connection 11, to a VDU 12 and/or a printer 13.

If the translation was to be done at the remote second location, the document would be sent by interface 7 via electronic connection 9 to a second computing means at the remote second location (not shown), which would communicate with code translator 10. Once the coding had been translated, the document could be accessed via VDU 12 or printer 13. In this case the transmission path would be from computer 4, via interface 7, electronic connection 9, second computing means (not shown), translator 10 to VDU 12 and/or printer 13.

Connections 9 and 14 are shown as dashed lines because they feature in only one of the two options discussed. Normally only one electronic connection 9 or 11, e.g. telephone or satellite link, would be used.

A further method of changing the mark-up coding would be to re-submit the required text, i.e. as shown at 25 on the right-hand side of VDU screen 20 in Figure 2, to computer 4. Using the coding facility 6, mark-up coding appropriate to the second equipment at the remote second location could be added to the text. It would then be possible to send the document via a direct electronic connection, e.g. telephone or satellite link 9, from interface 7 to the second computing means at the remote location (if required, but not shown) and then to VDU 12 and/or printer 13.

There is a further possibility for translating mark-up coding, which permits the information to be "translated" in such a way that the same document may be reproduced by the same system but displayed or printed to a different format. Thus, for example, if the same document was to be sent to two separate organisations A and B, but was required in a different format, this could be done simply by using different document profiles within the code dictionary i.e. by providing a dictionary which can interpret a mark-up code in two different ways. Assume that A wanted the document to the format shown at the right of Figure 2, then this could be printed directly via printer 8. If B's preferred format was, for example, to have all chapter numbers and titles underlined and commencing at the left-hand margin, then the operator would instruct the computer to use a different profile from within its own dictionary so that the mark-up codes would be interpreted according to B's requirements. That is, instead of interpreting code T2 as "title in capitals and centralised", the code would be interpreted as "title, underlined commencing at left-hand margin". Other codes could similarly bear different interpretations, according to different profiles.

According to an embodiment of the invention, there is provided an electronic text handling apparatus comprising:-

(i) means of generating an input, consisting of subject matter and generic, or mark-up coding, known as the source document;

(ii) computing means capable of processing said source document in accordance with the instructions interpreted from a generic code dictionary;

(iii) output means whereby said source document may be accessed to various formats according to the instructions interpreted by a particular profile selected from said generic code dictionary.

Preferably the means of generating an input would be a word processor or a VDU with editing facilities.

Preferably editing and formatting facilities would provide for display of the source document on one side of an A3 size VDU screen and the output subject matter on the other half of said A3 screen. It will be understood by those skilled in the art that the term 'access' means "being available", e.g. on a VDU screen, or as a printed document, or for further processing. It will also be understood that the 'dictionary' and the 'profiles' which form a part of it will have been programmed into the memory of the computing means.

According to an embodiment of the invention, there is provided an electronic text handling apparatus, comprising:-

(i) means of generating an input consisting of subject matter and a first set of generic coding known as the source document;

(ii) computing means capable of processing said source document in accordance with the instructions

interpreted from the first generic code dictionary;

(iii) translation means whereby said first set of generic coding may be wholly or selectively substituted by a second set of generic coding leaving said subject matter essentially unchanged;

(iv) output means whereby said source document may be accessed to various formats according to the instructions interpreted by a particular profile selected from the dictionary of said second set of generic coding.

According to an embodiment of the invention, there is provided an electronic text handling apparatus comprising:-

(i) means for generating an input consisting of subject matter being arranged in a first format;

(ii) means whereby said subject matter may be processed in which there is added a first set of generic coding resulting in the production of a source document;

(iii) computing means capable of processing said source document in accordance with the instructions interpreted by the first generic coding dictionary;

(iv) translation means whereby said first set of generic coding may be wholly or selectively substituted by a second set of generic coding leaving said subject matter essentially unchanged;

(v) output means whereby said source document may be accessed to various formats according to the instructions interpreted by a particular profile selected from the dictionary of said second set of generic coding.

The output means may either include VDU's or printers close to the source of creation of the input or it may include electronic connections, such as telephone, fibre optic, radio or satellite links, to

VDU's or printers at remote locations. The translation means may be situated either close to the source of creation of the input or at the remote location. Where the translation means is sited at the remote location, separate computing means will be required at said remote location.

CLAIMS

1. An electronic text handling apparatus, including:

input means operable so that a source document, entered into the apparatus, comprises a bare text, consisting of the subject-matter of the document free from formatting or output definition instructions, and mark-up codes, for specifying text format or defining text output presentation to be applied to the subject matter of the text, distinguished from the bare text,

processing means, including a store for holding a mark-up code dictionary defining interpretations of the mark-up codes for determining text format or text output definitions, operable to process the bare text in accordance with the mark-up codes as interpreted by the dictionary, and

output means, operable to access the source document with the bare text formatted according to the interpretated mark-up codes.

2. Apparatus as claimed in claim 1, further comprising:

editing means, whereby the mark-up codes, distinguished from the bare text, can be selectively edited separately from the bare text, to alter the formatting of the source document as assessed.

3. Apparatus as claimed in claim 1 or 2, comprising:

display means affording a dual display whereby the source document can be viewed in a first version, in which the bare text is displayed together with the mark-up codes, distinguished from the bare text in the display, and can be viewed in a second version, in which the source document is accessed and displayed as a formatted text, with the bare text formatted accord-

ing to the interpreted mark-up codes.

4. Apparatus as claimed in claim 1, 2 or 3, further comprising:

translation means, operable to detect the mark-up codes and to substitute for them, either wholly or selectively, different mark-up codes, according to a different mark up code dictionary, leaving the bare text substantially unchanged.

5. Apparatus as claimed in any preceding claim, wherein the dictionary contains a plurality of profiles, the profiles defining interpretations of the mark-up codes which are different as between one profile and another, operable so that different profiles may be selected for processing of the bare text in accordance with the mark-up codes.

6. Apparatus as claimed in any preceding claim, wherein the input means allow the bare text and the mark-up codes to be entered separately, for example the complete bare text being entered before entry of the mark-up codes.

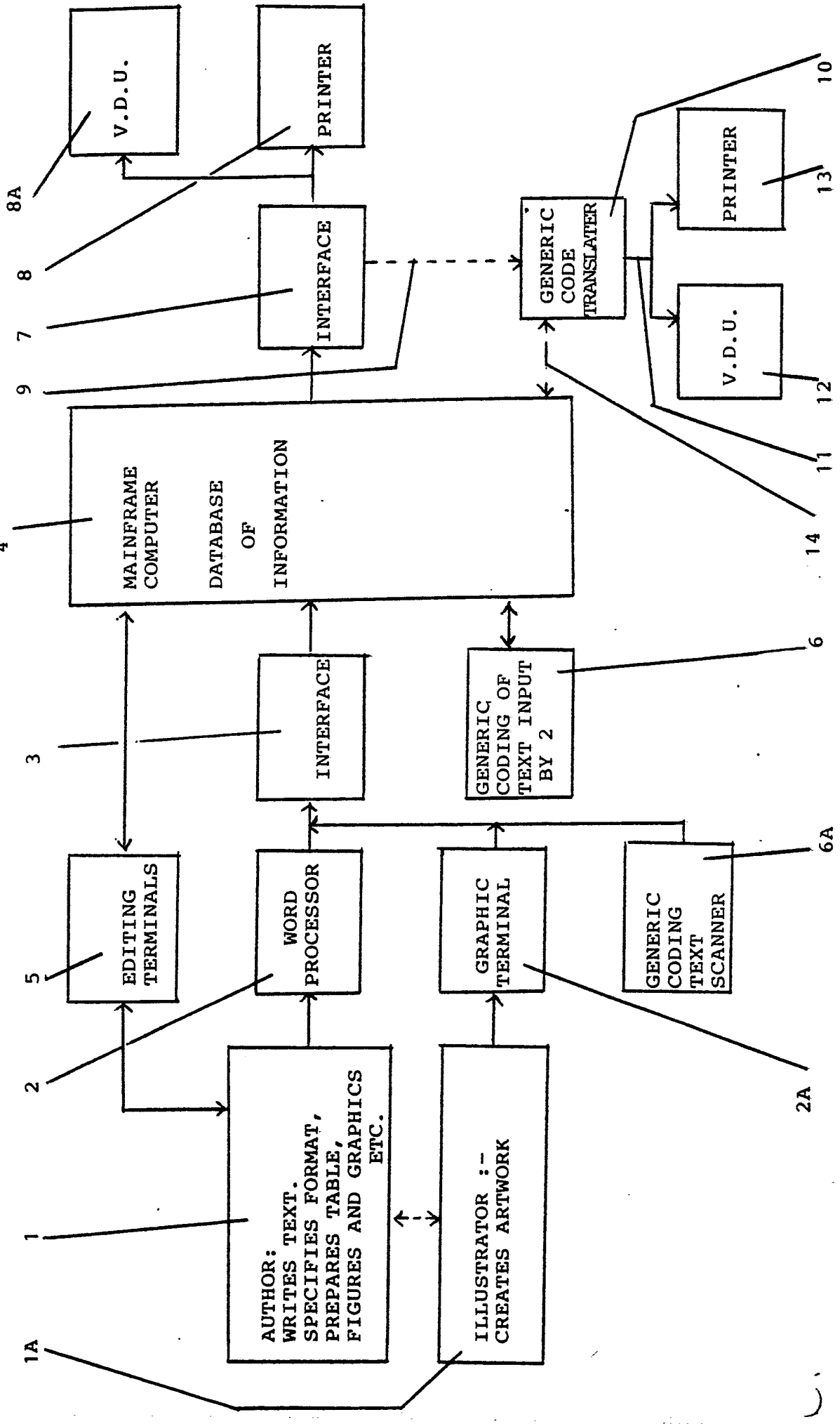

AUTHOR: WRITES TEXT. SPECIFIES FORMAT, PREPARES TABLE, FIGURES AND GRAPHICS ETC.
ILLUSTRATOR :- CREATES ARTWORK
EDITING TERMINALS
WORD PROCESSOR
GRAPHIC TERMINAL
GENERIC CODING TEXT SCANNER
INTERFACE
GENERIC CODING OF TEXT INPUT BY 2
MAINFRAME COMPUTER DATABASE OF INFORMATION
INTERFACE
GENERIC CODE TRANSLATER
V.D.U.
PRINTER
V.D.U.
PRINTER
1
1A
2
2A
3
4
5
6
6A
7
8
8A
9
10
11
12
13
14

FIG 1

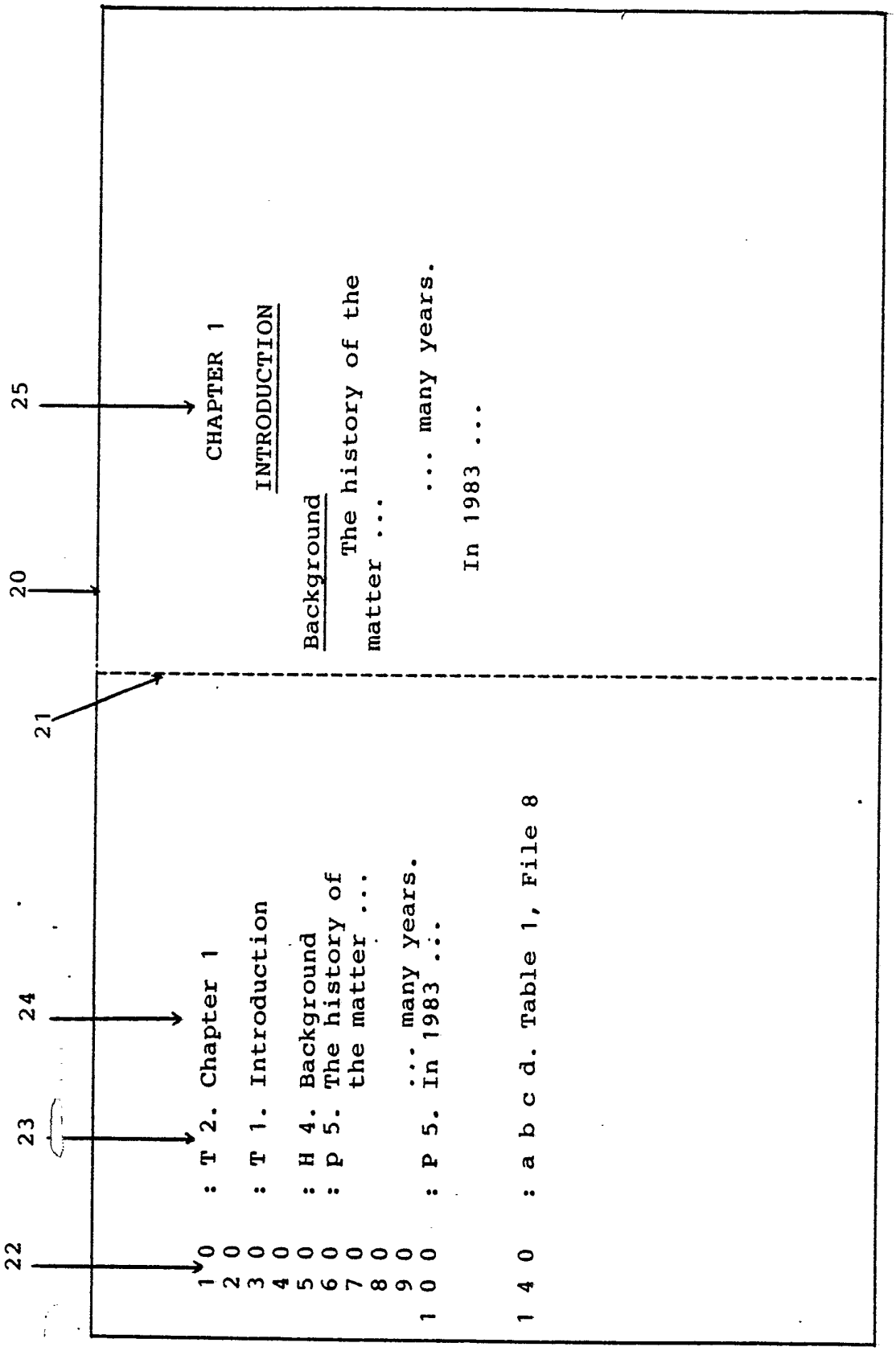

CHAPTER 1
INTRODUCTION
Background
The history of the matter ...
... many years.
In 1983 ...
10 : T 2. Chapter 1
20 : T 1. Introduction
30 : H 4. Background
40 : p 5. The history of the matter ...
50 ... many years.
60 : P 5. In 1983 ...
70 : a b c d. Table 1, File 8
80
90
100
140
20
21
22
23
24
25

FIG 2.