

Preparation for optical character recognition

By R. M. Paine*

This paper discusses the possible advantages of optical character recognition and matrix mark scanning for commercial organizations. Several available machines are mentioned and possible founts are considered in relation to the needs of Eastern Electricity. The problems of form layout, printing and paper quality are given attention, as well as the overall reject rates and the economics of O.C.R. The paper concludes with a review of actual developments and hopes for developments in optical scanning. This paper was first presented to the Glasgow Branch of the British Computer Society on 4th April, 1966.

The previous President of the British Computer Society, Sir Edward Playfair, said in his Presidential address in 1965 (Playfair, 1965): "In business applications the difficulty lies in the so far unsolved problem of the interface between the human brain and hand and the machine. No one can regard rows of girls reading figures and punching holes as an elegant solution of the problem in the computer age; but so long as computers have to deal with data produced by ordinary people, the problem must be faced. Progress is being made but all too slowly."

The author has been concerned to assess in Eastern Electricity the possible advantages and disadvantages of using machines to read reference numbers and amounts printed on documents by the computer line printer, and marks made on documents by meter readers. We have a very large Honeywell computer installation consisting of an 1800 computer and a 200 computer. (See Fig. 1.) The 1800 computer has 16,384 words, of 2 microsecond core store, each word being 12 numeric or 8 alphabetic characters, with 8 tape units working at 133,000 numeric digits a second; the computer is capable of running up to 8 programs at once. The Honeywell 200 has a core store of 4,096 characters, the access time to each character being 2 microseconds. Attached to the 200 are two tape units of 48,000 numeric digits a second, a card reader and one 900 lines per minute line printer; two other printers will be added later. We have 2 million consumers, and when they are eventually all on the computer system, the details required for consumer billing and accounting will occupy about 50 reels of magnetic tape. (See Grindrod 1964; Grindrod 1966.) On full load we will have approximately 100,000 input documents to deal with daily, some 72,000 of which we estimate could be read by some means of character recognition. Thus the reader will appreciate our interest in the subject and why we have carried out a few experiments. The documents we are concerned with for optical character recognition are meter reading slips and the stubs or counterfoils of the bills which are returned with the cash. We ruled out magnetic-ink character recognition since our line printer could not print magnetic characters and we were not prepared to imprint amounts in magnetic ink in our 150 shops or showrooms. Though we are a big concern, optical character recognition can be

used in much smaller organizations so I do not want you to think this is a technique for the "big league" only.

Desirable features of optical character recognition

The desirable features of O.C.R. as I see them are as follows.

- (1) Economically to replace key-punch operators from the point of view of:
 - (a) Cost of salaries and cost of equipment.
 - (b) Cost of space and the amount of space occupied.
 - (c) Burden of administration and supervision of staff of large punched card sections. (It has been estimated, however, that when our centralization is complete on all computer applications (not just Billing), we will still require 85 operators even with O.C.R. equipment.)
- (2) To increase the accuracy of data fed to the computer, since there is less chance of copying or punching errors.
- (3) To reduce the time to prepare and read data into the computer in order to avoid punching bottlenecks, and to lessen the elapsed time required to produce the final printed output—e.g. bills and meter reading slips—so as to have more chance of giving a 24 hour service to consumers.
- (4) To be able to use the same type fount as will be used for the rest of the document, and a fount which will be acceptable to consumers.
- (5) To have a high rate of correctly reading the characters and marks, and also a low rate of machine breakdown.
- (6) Not to increase significantly the cost of the stationery.
- (7) Not to increase significantly the operator effort on the printer or the maintenance attention on the printer, and to not reduce the speed of printing.

Founts

Turning first to the question of type founts—you will find there are plenty to choose from (see *Datamation*, 1966) such as the following.

* *Eastern Electricity, Ipswich, Suffolk.*



Fig. 1.—The computer room at Eastern Electricity with the Honeywell 1800 and 200 computers

- (a) IBM 1418 Fount—which is the same as that on the 1403 and 360 printers. This is a very acceptable fount to the consumer and could be used without worry on the rest of the document (*Data Processing*, 1963). (See Fig. 2.) IBM also have their later 1428 fount, which is stylized and similar to the A.S.A. “A” fount discussed in paragraph (c).
- (b) Farrington “self-check” fount. This is a more stylized fount, not so easy for consumers to read as 1418, so that one might consider in extreme cases having a different fount on the rest of the document which is to be read by people not by machines. (See Fig. 3.) “Self-check” by the way is an advertising name rather than one describing some particular inherent feature which causes self-checking.
- (c) The American Standards Association “A” fount—as used by Control Data and Standard Elektrik Lorenz (Mierzowski, 1966)—this is also stylized
- raising the same problems as the Farrington fount, but is an attempt to bring a standard fount to the business of O.C.R. (Orkild, 1963). It has been recommended by the International Standards Organization as one of their standard founts. (See Fig. 4.) Farrington’s new model optical character reader, the Series 3, 3010, can read the self-check fount and A.S.A. “A” fount, and IBM’s 1428 fount.
- (d) E.C.M.A. class “B” fount. A standard set by European manufacturers which is not stylized and seems likely to be very acceptable to consumers. I.C.T.’s Universal Document Transport will use this fount; no reader exists for delivery at the moment, though it is expected to be available early in 1967 on the I.C.T. 1900 series computers only. E.C.M.A. are also considering a lower-case standard. I.S.O. have also recommended this “B” fount, so although there are not any scanners

Fig. 2.—IBM 1418 fount

Fig. 3.—Farrington “self-check” fount

Some manufacturers do not read characters but coded bars to represent characters, normally numeric rather than alphabetic. Apparently coded bars are easier to distinguish than actual numbers or letters, and the speed of reading the documents may be faster. But they suffer

Fig. 4.—A.S.A. “A” fount

Fig. 5.—I.S.O. “B” fount

CONSUMER NAME AND ADDRESS

CONSUMER NO AND METER SEQUENCE NO					

二					二	二			二	二	二	
		二	二		二		二		二			
	二	二	二		二			二	二	二		

PREVIOUS 1	PREVIOUS 2	PREVIOUS 3	PREVIOUS 4	METER #	SEQ.	DAY
934	1020	1162	1290	04931768490	6	2
PREVIOUS 5	PREVIOUS 6	PREVIOUS 7	PREVIOUS 8	PRECEDING	EXPECTED HIGH	EXPECTED LOW
987	1920	1562	1092	18010	20010	18010

METER	0	1	2	3	4	5	6	7	8	9
DIAL 1										
10,000										
DIAL 2										
1000										
DIAL 3										
100										
DIAL 4										
10										
DIAL 5										
UNITS										

READING DAY	0	1	2	3	4	5	6	7	8	9
X10										
UNITS										

SPL. CHK. EST.

INITIALS _____

from two disadvantages:

- Two representatives of this type of O.C.R. are:

- HONEYWELL INSURANCE**

PAYMENT DUE NOTICE

YOUR POLICY NUMBER	PREMIUM DUE ON			CLASS	DISTRICT	BRANCH	PREMIUM
37100893M	MO.	DAY	YEAR				
	2	12	64	Q	417	392	15.06

GCB61637100893G081G01001506B

LOCAL OFFICE
ADDRESS

**RETURN THIS
NOTICE WITH YOUR
PAYMENT**



Mark scanning

The method of marking may be different according to the manufacturer, especially in regard to the guide lines to enable the human markers to place their pencil marks in the correct place.

An alternative to reading separate documents is to scan tallyrolls produced from cash registers. Such a system exists, for instance, with the N.C.R. optical reader, Model 420-2 reading tapes or rolls produced by N.C.R. accounting machines. This can mean that a by-product of the necessary cash receipting process can give you a character recognition input. We decided not to use this method because it would mean that the cashier in our showrooms had to insert a 12 digit consumer reference number on the keyboard of the cash register, which could lead to wrong numbers and would mean extra work for her. Also this system would not work for meter readings where we had separate documents, and we wanted to use only one type of optical scanner not two.

We decided to experiment to see what limitations, if any, O.C.R. would place on our stationery design. Straight away, we encountered a snag—the permitted sizes of form vary with each manufacturer, or in some cases depend on the number of characters in the line you want to read. For instance with IBM on the 1418 Mode I and II the document size was from $5\frac{7}{8}$ in. wide by $2\frac{3}{4}$ in. high, to $8\frac{3}{4}$ in. wide by $3\frac{3}{4}$ in. high. We had already predetermined to have 3 up printing for bills and 2 up printing for meter slips—this required the height of

the meter slip to be 4 in. and the width was $7\frac{1}{2}$ in.—which was too much for the IBM machine. Similarly with Auto-Lector: our bills were to be 5 in. wide by $3\frac{3}{4}$ in. high but with the 23 numeric characters we required to read, this width was too narrow for the Auto-Lector method and in addition the minimum size on the Auto-Lector is 4 in. \times 5 in. Thus working within our size of documents we could not experiment with all the methods.

The IBM 1418 code only works with an IBM 1401 or 360 computer attached, and interested as we were in O.C.R. we already had a Honeywell system and saw no reason to add an IBM computer to take over existing computer work, and to add a computer just for O.C.R. work made the cost of O.C.R. very high and uncompetitive with other machines. The I.C.T. "U.D.T." machine can be attached only to the I.C.T. 1900 series computers at the moment.

In the end the Chief Accountant of Eastern Electricity decided that we would test the Farrington machine. We were helped in this direction by the fact that, in the U.S.A., there existed Honeywell users with H.800 computers using Farrington O.C.R., so we were able to use their print barrels and equipment to carry out the experiments. Also Farrington equipment could read our size of documents. It was also possible with Farrington to use our check digit in the consumer number to check that this information had been read correctly from the document. We had previously decided not to introduce O.C.R. equipment until it was economic to do so in terms of replacing punch operators or saving the need to take on more punch operators. Thus we started our computer system using traditional key punch methods—which had an additional benefit in that we did not try to solve too many problems at once—and we were, therefore, not rushed with our O.C.R. tests.

Document layout

Most O.C.R. equipment requires that the horizontal length of the form should be greater than the vertical measurement—Auto-Lector is an exception to this—but this was required by Farrington equipment, and luckily it fitted into our stationery requirements with little difficulty. This may not be so, of course, for everybody's applications.

With Farrington it is necessary to designate one of the two horizontal edges as the feeding foot, and this edge should be as straight and un-mutilated as possible. With our bill counterfoils we had to make the feeding foot the bottom of the form since the top of the form may have been guillotined on a slant when issuing a machine receipt at the showrooms, or torn when detaching by hand the counterfoil from the top part of the bill along the perforation.

With Farrington the bottom of the line of characters to be read must be at least 2.187 in. from the feeding foot, which meant that for our forms the line had to be near the top of the form, though it could not be closer than 0.2 in. Our line was very near the top so as to clear other information we wanted to print on the form, and to

allow the desired distance of $3\frac{1}{4}$ in. for the matrix timing marks from the foot of the form. One useful point with Farrington is that one can overlap the positions of characters and marks so as to keep down the width of the forms—this is not always possible with other manufacturers. To fit in characters and a matrix did give some layout problems—for instance the reading zone was the line of characters (maximum number of characters is 60) $+ 0.3$ in. at each end, and a depth of 0.233 in., but since the height of the computer printed characters is 0.096 in., a certain amount of high or low printing can be tolerated. Around this reading zone had to be a clear zone of $+0.25$ in. at each end, and $+0.1$ in. above and below, in which no computer printing or background printing could appear except in an approved drop-out colour (see Farrington, 1963).

Within the line of characters one also has to be careful in layout, since if more than six spaces appear between fields, the timing of the scanner is upset.

The basic character set of the Farrington consists of the numerals 0 to 9, the letter H used as control symbol, and the minus sign. This gave us a problem since we wished to read amounts some of which would be credits which we were indicating on the bills by the CR symbol. This symbol could not be read by the Farrington basic set, and the full set including alpha would make the machine much dearer. We therefore had to decide whether or not to print credits on the counterfoil with a minus sign—thus risking confusing consumers with two methods of printing credits—one on the stub and one on the bill itself. We also had wished to use the plus sign for certain control purposes but were not able to do so because it was not in the basic set.

Colour printing problems

We had to decide in what colours to pre-print our forms. The line of characters would be printed by the computer as black as we could make it, and within the reading zone and clear zone no other colours must appear; but the colours used elsewhere on the form would make no difference to O.C.R. However, in the matrix area we wanted if possible to print brackets or squares to indicate to the meter readers where to put their pencil marks, and figures in them to show the value of the positions. These brackets could be blue or black since they were outside the area read by the matrix scanner. The timing marks for the matrix, however, had to be in black and since there had to be a very exact relationship between the timing marks and the brackets (if we had any) it was decided to print them both on the same run, so, therefore, the brackets would be black. The numerals within the brackets could not be black or blue since the matrix scanner might read them as information: red, however, was acceptable for the numerals since it is a "drop-out" colour not recognized by the photo diode used for mark scanning. After our experiments we were advised by Farrington that it would be better to have either no brackets or brackets in red since occasionally the black brackets could be read as a mark.

A later development of Farrington is an infra-red reading technique for matrix scanning so that if the timing marks are over-stamped with a non-dye ink (such as the mauve printing on our receipting machine or in cancellation stamps) the timing marks can still be clearly read. This is useful since with our timing marks at the top of the counterfoil they were very near where the machine stamped a receipt, and we were worried about what might happen with careless receipting.

There was a tradition in Eastern Electricity that our bills should be printed in blue and our final notices in red. This meant that on the front of the bill we had three colours

- (a) blue for the normal background
- (b) black for the timing marks and brackets
- (c) red for the numerals within the brackets.

The back of the bill had information printed on it, and although this was also in blue, because it was on the back the bill was regarded by the printers as a "4 colour job". The meter slips were only 3 colour jobs since there was no printing on the back. Now there are very few machines that can do a 4 colour job in continuous stationery and especially to the accuracy we required, so this colour aspect limited the number of printing firms who could do the job—and for our experiment only one firm printed the front and back of bills, though one other printed the front only, and three firms printed the meter slips.

Accuracy of printing

Very accurate printing is required and the tolerances are measured in ± 0.015 in. which is extremely difficult to achieve, especially when as on the matrix two colours in two separate runs have to be registered as accurately as this. This makes the printing job longer and consequently dearer. Character recognition by itself without the matrix would not be quite so difficult to preprint, and I wonder whether the matrix is necessary on the bill for part payments, which are a fairly low percentage. It might be easier and cheaper to mark a special circle for part payments which would cause the scanner to reject the counterfoil, and we could then punch up the information on that stub. The matrix is needed on the meter slip but there are only three colours involved here, which makes printing easier, and we might accept the printing of the background in black rather than blue, since it is an internal document, making it a 2 colour job. Also we might remove the brackets entirely so that there is a little more tolerance in printing the red numerals on the second run.

In our experiments about 50% of the forms supplied by 2 printers had timing marks which were either $3\frac{7}{32}$ in. or $3\frac{3}{16}$ in. from the feeding foot instead of $3\frac{1}{4}$ in., and this caused a certain number of rejections and bad reads. Some of the documents had also the measurements, from the top of the timing marks to the centre line dimensions of the red numerals in the matrix, outside specification by 0.078 in.

There is no doubt that O.C.R. does impose on the printers of forms very stringent specifications, but once these have been realized and accepted by British printers they should be perfectly capable of producing the quality required.

Quality of paper

Most printing firms quoted for 21 lb paper but these papers in some cases had "inclusions" or specks of wood in them and were not quite a good enough quality or substance. Farrington were happier when in one test we used 24 lb paper, which was called "dust free" without inclusions. I believe that Dividend Loan Paper is of the right quality. Because of the poor quality of some of the paper, which was rather porous, the density of the pre-printed timing marks and the computer print itself were below standard and contained voids and extraneous ink. These were not always apparent to the naked eye.

But by proper choice of paper and regular attention daily to the computer printer and frequent changes of the carbon ribbon, it should be possible to produce satisfactory printing. We did not have to reduce the speed of printing but if a slower speed is feasible in an installation I think this would produce good results more consistently.

For normal printing of bills—without O.C.R.—we use 18 lb paper and the extra costs involved in going to 24 lb and the special preprinting are about 3s. 0d. per 1,000 bills. We will print ultimately about 8 million bills a year, so this extra expense must be taken into account.

Speed of scanning

The speed of reading by Farrington depends on the size of paper and we expected to be able to read about 250 to 300 documents a minute. A similar speed of 300 to 400 documents a minute would be obtained with the IBM 1418 or the Auto-Lector depending on the size of the documents. Honeywell's bar scanner was planned to read at least twice as fast, and the S.E.L. O.D.S.2 scanner read about 600 documents a minute. The R.C.A. Videoscan is reported to read from 750 to 1,500 documents a minute depending on size of document. With 72,000 documents to read daily the Farrington scanner would still mean a load of 4 to 5 hours a day, allowing for handling, so there still could be a bottleneck in trying to read all the day's input by midday, in order to process the transactions and produce bills that could be despatched that same evening.

The method of output from the Farrington scanner allowed a wide range of choice and also presented difficulties. If used off-line one could produce paper tape, punched cards or magnetic tape. Paper tape or punched cards would mean reducing the speed of dealing with documents to that of the speed of the punch—and this was a speed reduction we could not afford—especially as we would then have to convert to magnetic tape before processing the input on the H.1800.

We therefore wanted to go direct to magnetic tape—

but unfortunately all the character scanners on the market produced $\frac{1}{2}$ in. magnetic tape off-line and we required $\frac{3}{4}$ in. tape for our Honeywell system. It looked, therefore, if we wanted to use the Farrington scanner or any other non-Honeywell scanner, we might have to have it on-line to the 200 or 1800 computer producing a $\frac{3}{4}$ in. magnetic tape—and we have then to take into consideration the problems of keeping a fairly slow speed device on-line to a computer for 5 hours or more a day. It probably would not slow down our other jobs significantly, but space for a small program must be left in core store thus prohibiting any of our programs using the full core, and it would also mean one tape unit almost permanently linked to the scanner, which would cut down the number of tape units available for our processing programs. A better alternative solution is to have the scanner off-line producing $\frac{1}{2}$ in. magnetic tape, and then on the H.200 to have a $\frac{1}{2}$ in. tape deck so as to transfer the data from $\frac{1}{2}$ in. to $\frac{3}{4}$ in. tape for processing on the H.1800. This would also give us the advantage of being able to read other industry-compatible $\frac{1}{2}$ in. tapes on the H.200.

Accuracy of scanning

The accuracy of scanning seems to be shown in two main ways.

- (a) The percentage of errors—i.e. digits said by the scanner to be one character but which were in fact another character. Very few figures are ever quoted for this since you do not know at the time of reading that it is an error, but the percentage appears to be very low indeed.
- (b) The rejection rate—i.e. documents rejected by the scanner because the machine could not decide what the character was, or because the documents were too badly crumpled or mutilated to be read. With the Farrington and other scanners this rate seemed to vary with the installation from about $\frac{1}{2}$ % to 4 % of the documents. This figure depends to some extent on whether it is the practice of the installation to feed all the documents to the scanner, or remove some manually in a prior stage because one could see they were too badly damaged or stained to go through the scanner.

In our experiments the combined error and reject rate for meter slips was around 5%. Some of the documents were purposely mutilated to see what would happen—but these rejects have been excluded from the 5% figure. I do not think this result can be used for or against any line of policy. I feel that given more time to check the standard of computer printing and using only the best paper samples, we could have improved on this performance. We were, however, carrying out our experiments in America, not in our own installation, so it was difficult to take the time to make all the adjustments we would have liked.

In Japan the Kansai Electric Power Company using 2 Farrington machines processed 814,036 documents in a

week and had only 11,808 rejected. This is a percentage of 1.4 % and these forms included characters and marks and had been filled in by 1,600 different meter readers. The overall speed of processing, including handling, was 240 documents a minute. So obviously it can be done successfully!

I/S Datacentralen in Copenhagen are also successfully using an optical scanner for processing motor vehicle licences, as I saw on a recent visit (Stangegard, 1966), and the public are not worried by the Self-Check fount.

Cost of equipment

A Farrington scanner with punched card output costs about £40,000, and with $\frac{1}{2}$ in. magnetic tape output is about £54,000. Similar prices exist for other scanners, though for the IBM 1418 and the I.C.T. U.D.T. you also have to take into account the cost of the associated computer. The Auto-Lector is cheaper, its cost being around £20,000 to £30,000. I estimate that a Farrington scanner could do the work of 40 to 45 key punch operators, and taking into account their salaries and the rental of punches and verifiers one could spend about £27,000 a year on punching—thus the scanner would soon become economical, even though in this quick comparison I have left out other costs associated with the scanner. At a rough estimate I would say that if with a scanner you can save 15 to 17 girls, then in 4 or 5 years you would have more than broken even. One of the problems is knowing when to introduce the equipment, since your input builds up gradually over the months as you convert the current system to the computer, and you cannot wait until you have 45 girls on the work that could be done by the scanner, and then replace the lot. It is difficult enough to recruit the girls in the first place, and one must try to bring in the machines at a point where it is economical to do so, but at the same time avoid staff problems such as redundancies. Perhaps the best time to introduce a scanner is when it can take on the work of about 15 to 20 operators, and these operators can be used on other punching or moved into clerical work so as to reduce redundancies.

Other developments

Optical Character Recognition will I believe bring large benefits to Eastern Electricity, but I feel something more is required, to reach real automatic input. We have to move our documents to be read from at least thirty offices to Ipswich, which is 90 miles away from some of these offices. Time is lost in the post and even if the documents are available in the afternoon at the local offices, it is the following morning before a start can be made on feeding the documents into the computer. What I would like to see is a cheap system of character recognition that could be used at the local offices, and for the information read from the documents to be transmitted by telecommunications direct into the computer during the evening and night.

A start has been made on such a system in the U.S.A. where Honeywell are supplying bar code readers at 900

local offices of the Metropolitan Life Assurance Company, to send information by telephone to an H.1800 computer in New York. The O.C.R. machines actually read the documents fairly slowly and punch the information into paper tape which is then used for later transmission. Only a few offices have so far been attached to the head office in this way, but it is a very interesting development (Diebold, 1965).

In the U.S.A. I found firms that were typing information and then using O.C.R. to convert it to magnetic tape, instead of punching and verifying cards, in order to set up large files of data or to update their master files. The advantages claimed for typing and O.C.R. over the traditional method of punching and verifying, are:

- (a) Typing is more accurate.
- (b) Typing is faster.
- (c) The verifying operation is replaced by spot checks and proof reading which is said to be faster and more accurate.
- (d) Although an individual typist may be paid more than a punched card operator, they are easier to obtain or to train, and, therefore, the overall cost is cheaper.

I would not like to offer an opinion on this, but it is always worthwhile questioning existing methods and I believe a computer bureau in London is going to offer this service for file conversion.

Time and Life Corporation in Chicago use this method to introduce new subscribers to their magnetic tape files (see Fig. 8) and a service bureau firm in New York—New Era Data Systems—offer a service of this kind to their clients to set up mailing lists, stock holder records, car licence registrations, personnel records etc. The O.C.R. equipment they have used for some years is not a document reader but a Farrington page reader which can read many lines on a page—similar equipment is sold by C.D.C. and IBM.

I saw a very interesting O.C.R. machine made by Philco Corporation, a division of the Ford Motor Company, which could read several different types of founts at a speed of 2,000 characters a second. It would be useful to a large organization to whom documents were sent from and prepared at many different sources, over which they had no control as to the type of fount to be used. The founts can be mixed in a batch of documents, or even on one form. This was, however, a very costly machine priced at about £150,000 and I cannot see a large market for it, though New Era Data Systems have installed one in their service bureau.

References

- FARRINGTON (1963). "Farrington Bradma, Electronic Reading Machine, Model 10 Series, Specification."
 STANGEGARD, O. (1965). "Optical Processing of Tax Forms", *International Business Equipment*, April 1966.
 PLAYFAIR, E. (1965). "Presidential Address", *The Computer Bulletin*, Vol. 9, No. 3, December 1965.
 GRINDROD, F. S. (1964). "A Computer for Eastern Electricity", *Local Government Finance*, May 1964, pp. 151–153.
 GRINDROD, F. S. (1966). "Installing a Centralised Computer", *Local Government Finance*, April 1966, pp. 127–129.
 DATA PROCESSING (1963). "Optical Reading Machine—The IBM 1418", *Data Processing*, July–August 1963.

```

11 0088 MRS ADA KUHLEN
8641 CHEVEL
8641 CHEVY CHASE
0642LA MESA CA
524 402493 2 9 104 4 0675

10 0088 COL CARL P SCHNEIDER
BLDG-GRADS
UNIV-CINCINNATI
0642CIN 21
524 402121 297
524 402121 2 9 104 4 0675

00 0088 PUBLIC LIBRARY
BROWN1
0642BROWNS VALLEY MN
524 402119 2 9 156 1 1875

11 0070 MRS JOSEPH HINCHEY
137 TWIN OAK DR
0642LEVITTOWN PA
524 401438 2 9 104 1 1275

11 0070 MRS ROGET
11 0070 MRS ROBERT HUMMER
    
```

Fig. 8.—Specimen input to a Farrington page reader. (The machine will ignore corrections)

One organization that will have such multiple fount problems is the G.P.O. when they introduce their GIRO system in Britain. As I understand it, people with deposits in the GIRO will be able to take any bill including gas, electricity and rates into the post office, make out the necessary form and ask the post office to pay the bill by transfer from their account, or pay cash over the counter to settle the bill. Now public utilities like to be told who has paid and how much—and the G.P.O. will want to know as well. The public utilities will want to do this increasingly by O.C.R. on the payment stubs—but there is no standard fount or system among the utilities. The G.P.O. will also want to read this information by O.C.R. but they may use different systems from Gas and Electricity Boards. If the GIRO really catches on in this country I can see a great impetus being given to O.C.R., and also a strong move to some form of standardization of fount and system.

I would like to conclude with a hope that O.C.R. difficulties will be overcome and that we will see some very useful and cheap developments in the future, and perhaps some of the things that at the moment seem rather "way out" may yet come to our aid to provide Sir Edward Playfair's "elegant" solution. I would like to thank the Chief Accountant of Eastern Electricity for permission to publish this paper.

- ORKILD, C. O. (1963). "Standardisation—Bulletin on Optical Character Recognition", *The Interpreter*, November 1963.
DATAMATION (1966). "Table of Character Recognition Machines", *Datamation*, February 1966, pp. 48 and 49.
DIEBOLD, J. (1965). "In the Midst of Metropolitan Life", *Data and Control*, January 1965.
MIERZOWSKI, P. (1966). "Optischer Dokumentsortierer OD52", *SEL-Nachr.*, Volume 14 (1966), No. 1, pp. 39–44.

Correspondence

Systematics

To the Editor,
The Computer Journal.

Sir,

C. B. B. Grindley in his article on Systematics (this *Journal*, Vol. 9, No. 2, page 124) suggests that the systems analyst need have no computer knowledge since he is concerned solely with expressing the relationship between the data fed into a computer system and the information to be produced by it.

I believe this to be a dangerous assumption. In practice the design and even the objectives of a system must be greatly influenced by hardware and software considerations if the system is to be practical and economic. Thus the systems designer must draw heavily on his experience of the way a computer actually processes data.

Rather than attempting to place the different aspects of commercial computer work into water-tight compartments, I prefer to look at the relationship between them. It seems to me that proficiency in basic programming techniques is necessary before a person is qualified to undertake the technical specification of a large commercial system involving the manipulation of many files and the design of an inter-related set of programs. Similarly, it is necessary to assess the technical implications of a system whilst deciding its objectives and the way it will operate in a particular environment.

A reason that is often advanced for making a clear-cut distinction between the programmer and the systems analyst is that they are two essentially different types of people since one deals with introverted technical problems whilst the other is concerned with communicating ideas to the outside world. By the same token every profession should be made up of two classes of people. One to sit in the office and talk to the clients and the other to perform the irksome task of actually applying the tools of the trade.

If this happened then we in data processing would not be alone in having a communication problem.

Yours faithfully,

D. A. BURBRIDGE

85 Gathurst Road,
Orrell Post,
Wigan,
Lancashire.
8 August 1966.

To the Editor,
The Computer Journal.

Sir,

I was most interested, and I am sure many others were, by Mr. C. B. B. Grindley's definitions of *Programming* and *Systems Analysis* in the last issue of this *Journal* (August 1966, page 126). His *Programming* was succinct and accurate, but his *Systems Analysis* was restricted solely to the part of the job which involves an "interface" with the programming function.

May I suggest the following extension to his definition. I also repeat his *Programming*.

Systems Analysis: Analysing the needs of a potential computer user and designing an automated system utilizing a computer which represents the optimum compromise between the user's needs and the capabilities of a particular computer system; expressing the relationship between the data to be fed into the computer system and the information to be produced by it.

Programming: Expressing the relationship between the data fed into a computer system and the information to be produced by it in a manner which is efficient in terms of the capabilities of that system and in a manner which can be interpreted by that system.

Yours sincerely,

A. ROBERT BROWN

36, Greencroft Gardens,
London, N.W.6.
19 August 1966.

To the Editor,
The Computer Journal.

Sir,

Efforts, such as those of Mr. Grindley, to improve on the formal presentation of analysis should certainly be encouraged. Unfortunately Mr. Grindley's paper contains a number of ridiculous remarks which we are surprised to find in *The Computer Journal*.

Surely the only people who can seriously believe that "the programmer is solely concerned with efficiency problems involved in processing" are those analysts who feel that they "require no computer knowledge" and others who are similarly ignorant of this field.

These quotations are, however, only an extreme form of a common fallacy. The belief that Systems Analysis has "emerged as a prior and separate task from programming" arose, in our experience, from the heavy demands on programming resources and the fact that the scope of earlier computers could be appreciated by the intelligent layman who could organize many applications using teenage coders. One of the consequences of the new series of machines is to emphasize the importance of programming. For example, since the configuration is now so dependent on the requirements of the operating system, even the choice of machine demands access to sound programming knowledge.

At a time when the Society is investigating professional qualifications, it becomes urgently necessary to define, or at least describe, the relevant computing skills. We regard programming as all the work directly concerned with the computer implementation of an identified application, together with advice on the suitability of an application in terms of available computing facilities. We feel that coding has

(Continued on p. 237)