Status report on IBM document reader

Webb, E. L. R.

Publisher's version / Version de l'éditeur:
https://doi.org/10.4224/21273574
Report (National Research Council of Canada. Radio and Electrical Engineering Division : ERB); no. ERB-513, 1959-05

[redacted]

ERB-513

LIMITED

FOR CANADIAN GOVERNMENT
USE ONLY

NATIONAL RESEARCH COUNCIL OF CANADA
RADIO AND ELECTRICAL ENGINEERING DIVISION

STATUS REPORT ON IBM DOCUMENT READER

E. L. R. WEBB

OTTAWA

MAY 1959

Limited

STATUS REPORT ON IBM DOCUMENT READER

- E.L.R. Webb -

A visit was made on March 20-25, 1959 to the Glendale Product Development Laboratory of IBM at Endicott, N.Y., to check on progress of the document reader intended for the 1961 Canadian Census. In addition to a larger sample of test documents for statistical purposes, a small batch of miscellaneous sheets was run through the machine to test for several specific ways in which errors might occur. The selection of markings was by no means exhaustive but included some of the aberrations in pencil marking that might be expected in practice.

The permissible marking positions on the test sheet form a matrix of 35 rows and 32 columns on the lower half of each face of the 14" × 16" paper. The element or lattice points on opposite sides of the sheet are staggered half a row and half a column to minimize cross-talk; however, grossly oversize marks could cause trouble. For convenience in logical handling and print-out the columns are partitioned into 5-row groups. Empty groups result in a print out of the symbol 8, whereas groups containing a single mark yield one of the five symbols "—", "(blank)", "4", "2", "1", depending on the location of the mark in the group. Groups with two or more marks would give other symbols according to the IBM tape code.

The 64 columns on both sides of the sheet are read in parallel by 64 silicon solar cell channels. Paper motion being in the direction of columns, the waveform in each channel represents the profile of all elements in a given column. Gating circuits distribute the elements, row by row, into appropriate buffer storage units, whence the information is later transferred to magnetic tape. In the NRC test set a single column is scanned repeatedly on a drum turning at 150 rpm, which is fast enough to give a useful picture on a medium or long persistent CRT screen.

The small batch of test sheets consisted of 14 sheets numbered G-O to G-13 for identification and was run through three times for redundancy.

Document G-O was blank and served as "leader", as well as to check for "false alarms" or over-sensitivity. Sheets G-1 and G-6 were marked in a standard test pattern with a mark-sense pencil using normal pressure and served both as "carbon paper" for transfer tests and also as normal documents to check for "missed targets" or gross under-sensitivity.

Documents G-2, 3, 8, 9, 10, 11, and 12 were either real or simulated carbon transfers with the critical marks located in the interstitial positions of the lattice or matrix. In nearly all cases the IBM reader successfully avoided reading these marks which were in nominally blind regions; however, the front face of

G-10 reported a significant number of marks. Upon visual examination they appeared to be elongated marks extending into the active columns, but most were well centered between rows and should have been excluded by the electronic gates. It would appear that the effective gate openings are longer than we had been led to believe.

To test for weak mark sensitivity, G-4 was marked in the same manner as G-1 and G-6 but with reduced pencil pressure. The resulting print-out was discouraging at first, but later proved consistent with the then current condition of the reader as determined from a detailed examination of sheets G-5 and G-13 using the NRC test set. It turned out that the front side of the machine had at least adequate sensitivity but the rear side did not, and was somewhat ragged. It is now felt that failure to read weak marks should not be a serious source of error.

The unlikely but possible source of error by "punch through" was investigated on sheet G-7 by making heavy marks on the reverse side but in the correct position for reading from the front. In general the paper has sufficient opacity and the few marks reported corresponded to noticeable mechanical distortion of the paper.

The two remaining documents G-5 and G-13 were given detailed examination by means of the NRC test set which simulates closely the reading function of the IBM machine. However, the test set accommodates only a narrow width of paper so it was necessary to slice the documents into strips, thereby destroying some of the information on one face of the paper, but enough data was obtainable from a single side of each sheet. For a number of reasons, which will be apparent, the rear face of G-5 and the front face of G-13 were selected for examination.

Document G-5 was marked with a hard (Venus 2H) pencil so as to produce marks of marginal density. From the print-out it was conjectured that the general sensitivity of the front half of the IBM reader was adequate, but that the rear half was in difficulty.

Using the small printed dot at the end of each column as a reference, the critical or triggering level of the NRC test set was adjusted to match the sensitivity specified by IBM. Then columns 33 to 64 were viewed individually and each pencil mark rated according to whether or not it should have registered. (In fact about 10% were too weak to trigger).

The performance of each solar cell channel was assessed by counting the agreements between the test set rating and the print-out data for the three runs. Since there were seven marking positions used per column, the maximum score was 21, which was attained by only seven channels. Three more scored 20, corresponding to less than one error per pass. Ten channels had scores of 19, 18, or 17 corresponding to one error and the remaining twelve scored 16 or less — including four real duds with scores of 5 or less. The responsibility for this high error rate must lie with the machine because due credit was allowed for weak pencil marks. The cause may have been reduced light output from the rear exciter lamp — common to channels 33 to 64 — for it appears unlikely that a large number of sensitivity failures should occur selectively on one side of the machine. (The front side showed very few, if any, such failures.)

Document G-13 was an erasure test. It was first marked normally, with a mark sense pencil on the front face and a hard (Venus 2H) pencil on the rear; then every even-numbered column was erased. Again using the column-end dot as reference, columns 1 to 32 were examined and rated as before. Also the channel performance was assessed by correlation between test set data and the IBM print-out. All sixteen of the unerased columns had perfect scores, as did eight of the erased columns. Here, of course, success consists of ignoring any residue left after erasure. Another seven channels, with scores of 20, 19, or 18 may have been slightly over-sensitive because in no case did an erased mark trigger the test set. The remaining channel (No. 28) scored 11 and evidently was "trigger happy".

The general lack of sensitivity in the rear half of the document reader reduced the opportunity to extend the erasure test. However, of the ten good channels in the group 33 to 64, six corresponded to erased columns, and in none of these did the print-out data indicate false alarms. The four good channels reading unerased columns were well behaved, but of the remaining twenty-two channels nothing could be said. On the whole it would appear that erasures should not be a serious source of error if done with reasonable care.

In summary, it appears that as of March 25th, only the front and about a third of the rear sides of documents were being read reliably. This condition could, of course, change radically with variations in parameters such as channel gain, trigger level, exciter lamp voltage or solar cell sensitivity. In addition to faulty readings, the strictly mechanical functions of document feeding and re-stacking were giving intermittent failures which, in the long run, could be very troublesome. Another difficulty encountered on a previous occasion, namely failures in the buffer storage area, appears to have been corrected.