

Griffith

COMPANY CONFIDENTIAL

PROJECT STRETCH

FILE MEMORANDUM #33

SUBJECT: Tentative Specifications for a Printed Document Reader

BY: W. Buchholz

DATE: April 18, 1956

The following specifications are for a reader which will be the primary document input for LINK (the input-output system of STRETCH). The documents will contain only human-readable characters.

Additional input for punched cards or other kinds of coded documents will be provided only as needed to communicate with other systems using such coding.

These specifications represent what appears to be a reasonable goal for a document reader development effort. The LINK system, however, will not be limited to the performance and features contained herein. As experience is gained, the specifications may be revised to take advantage of any higher performance or new desirable features found obtainable without substantially increasing the complexity or sacrificing reliability.

1. Scope of Reader

The reader should include

- an automatic document feed,
- a scanner, and
- circuits to reduce the scanned information to standard signals which are independent of the programmed recognition logic.

The reader should not include circuits or programming means for the recognition of specific characters. Recognition will be done by stored program in the LINK computer.

The dividing line should be determined by delegating to the computer those logical functions which must be changed when going from one alphabet or type style or document arrangement to another. All functions which can remain fixed from one application to the next should be built into the reader.

April 18, 1956

In the design of the reader strong emphasis must be placed on mechanical simplicity and reliability.

2. Document Feed

The feed must accommodate both paper and card documents of sizes and thicknesses found in common use. They should include sheets as large as 8 1/2 x 11 inches and stubs as small as 2 x 3 inches. Extremely large and extremely small sizes will be considered to be special cases which may require a special adaptation of the design.

Documents in fairly good shape should feed automatically, but provisions must be made for hand-feeding individual documents.

3. Character Styles

All alphanumeric characters printed by standard IBM printers should be recognized.

The type styles should include:

- (a) the commonly used printer's type fonts,
- (b) the commonly used typewriter type fonts, and
- (c) characters lettered by hand with reasonable care.

Since recognition of specific characters is to be done by computer programming, different programs can be called in to handle various ranges of type styles.

4. Registration

It must be possible to skip at high speed to any line on the document and to any word in the line under program control. Typically, only a small amount of the information printed on a document needs to be read, and the time spent passing over large unwanted areas of a document must be kept low.

Registration tolerances within a line must be broad enough to permit variations unavoidably encountered with typewritten and hand-lettered material. Some form of self-aligning scanner is indicated. Such a scanner would also simplify the design of the feed.

5. Speed

In general, the higher the speed, the more attractive the reader will be. The limit on speed is expected to be set by design limitations

of the reader, not by the LINK system. On the other hand, high speed is not essential to the initial success of the reader, and it is important not to sacrifice reliability and flexibility to get higher speeds.

The following nominal speeds might be considered as a starting point:

Reading - 500 characters/second
Feeding - 250 documents/minute

Feeding is assumed to be intermittent, with the document being stopped for reading. The nominal feeding speed applies when no reading takes place. Reading time is additional, so that the net feeding speed decreases as the amount of information to be read increases.

The feed should have multiple-point clutching, or the equivalent, so that the drop in feeding speed is kept to a minimum. It should also be possible to increase the feeding rate for smaller documents, so long as the maximum linear paper speed remains constant.

6. Programmed Logic

As already mentioned, the recognition logic will be programmed in the computer. Programming will also be used for:

checking information on the document,
correcting errors on the document,
field definition,
zero insertion, and
symbol deletion.

7. Code Reading

For most applications there will be no need to read other than human-readable characters, and hence no special provisions will be made to read coded marks.

It is expected, though, that by suitable programs, but using the same feed and scanner, it will be possible to recognize many kinds of printed coded marks and punched holes encountered in documents designed for other systems.

8. Optical Scanning

Optical scanning of regular black ink, or of other visually dark marking, is required for general-purpose document reading. It permits the use of standard printer's inks, typewriter ribbons, carbons, pencils, and pens. This is an important requirement for source recording where special devices or materials are not easily available.

An even more basic reason is that visible marks can be visually inspected by clerks for reading quality, even after erasure and correction. With optical scanning the criterion for quality is similar for human or machine reading. Fluorescent or magnetic inks have been proposed for protection against obliteration. Here the standards of legibility are quite different and clerks have no way of inspecting the writing for machine readability without access to special devices. Fluorescent or magnetic inks would continue some aspects of the dual standards of recording which have proved troublesome with punched card documents.

An optical scanning and recognition device will protect against errors due to obliteration so long as it rejects all unreadable characters and does not substitute one character for another. Experience to date has shown that substitution can be prevented. The addition of redundant characters with programmed error correction, where practical, can be used to decrease the rate of rejection.

9. General Comments

The general-purpose reader will not need multiple stackers for the purpose of sorting. Document sorting requirements are rare enough so that special equipment is indicated.

No provisions will be made for printing identifying information in the reader. It is expected that any such validation printing will be done in a separate device, and the reader can subsequently pick up the identification.