Title
Emanuel Goldberg, electronic document retrieval, and Vannevar Bush's Memex

Journal
Journal of the American Society for Information Science, 43(4)

ISSN
0002-8231

Author
Buckland, MK

Publication Date
1992

DOI
10.1002/(SICI)1097-4571(199205)43:4<284::AID-ASI3>3.0.CO;2-0

Preprint of article in the *Journal of the American Society for Information Science* 43, no. 4 (May 1992): 284-294, published for the American Society for Information Science by Wiley and available online to ASIS members and other registered users at http://www.interscience.wiley.com/. Text vary slightly from the published version.

Emanuel Goldberg, Electronic Document Retrieval, And Vannevar Bush's Memex

Michael K. Buckland
School of Library and Information Studies, University of California at Berkeley, Berkeley, CA 94720.

Vannevar Bush's famous paper “As We May Think” (1945) described an imaginary information retrieval machine, the Memex. The Memex is usually viewed, unhistorically, in relation to subsequent developments using digital computers. This paper attempts to reconstruct the little-known background of information retrieval in and before 1939 when “As We May Think” was originally written. The Memex was based on Bush's work during 1938-1940 developing an improved photoelectric microfilm selector, an electronic retrieval technology pioneered by Emanuel Goldberg of Zeiss Ikon, Dresden, in the 1920s. Visionary statements by Paul Otlet (1934) and Walter Schürmeyer (1935) and the development of electronic document retrieval technology before Bush are examined.

Introduction

Vannevar Bush's description in 1945 of an imaginary information machine, the “Memex”, is constantly viewed and cited in relation to *subsequent* developments in computing, information retrieval, and hypertext. Seen this way Bush and his Memex appear as strikingly original and visionary. Little attention, in contrast, has been paid to Bush's Memex in relation to its *own* context: the visions and technological developments of information retrieval in the 1930s. Bush and the Memex when seen in context look rather different from the more familiar view of them without their context. This paper examines the technological background of the Memex and of other visions of that period, with special reference to Emanuel Goldberg.

“As We May Think”

Published in 1945 in the *Atlantic Monthly* Vannevar Bush's description of an imaginary personal information machine in his article “As We May Think” had an immediate impact (Bush, 1945a). *Life* published an illustrated version entitled “A top US scientist foresees a possible future in which man-made machines will start to think” (Bush, 1945b). *Time* summarized it briefly under the heading “A Machine That Thinks,” with a photo of Bush bearing the possibly ambiguous caption “Prof Bush: Just turn the crank” (Anon., 1945). Bush reprinted his article in his *Endless horizons* (1946, 16-32) and wrote about it further in two later books: *Science is Not Enough* (1967, 75-101) and *Pieces of the Action* (1970, 190-192).

“As We May Think” has been constantly cited ever since, a development that has been analyzed by Smith (1981), who noted that the article has been used as a symbol for a number of different concepts. However, references to it often have little substance. The paper has become a fashionable icon of modern information science, typically used as a convenient point of departure, or as an invocation of respectability. Bush has even been hailed as a “Father of Information Science” (Lilley & Rice, 1989).

It is difficult, now, when reading “As We May Think”, not to think of how Bush's suggestions are coming to be realized through the increasing power and versatility of modern digital computers. Viewed in relation to developments in information systems *since* 1945, it is easy to see Bush's Memex as a beacon pointing out what should follow. But this is an incomplete and unhistorical perspective. Although published in 1945, the paper was originally written in 1939 (Nyce & Kahn, 1989) and it had nothing to do either with digital computers, which were only then beginning to be invented, or with the analog computers on which Bush himself had worked.

The Memex

The Memex itself is an imaginary personal information system designed around a sophisticated microfilm reader and with more functionality than any microfilm reader ever built. Bush discusses it in relation to the needs of scholarship rather than business records.
The imaginary Memex has the following components:

1) A collection of documents on microform. In this Bush is following a long tradition. The idea of providing convenient copies of documents on microfilm dates at least from 1853 (Stevens, 1968: 363) and had become a common notion. For example, Paul Otlet (1868-1944), the Belgian documentalist, and Robert B. Goldschmidt (1877-1935), a Belgian inventor, had proposed standardized microfiche in 1906 (Goldschmidt & Otlet, 1906; Otlet, 1990, 87-95) and a portable microform library in 1925 (Goldschmidt & Otlet, 1925; Otlet, 1990, 204-210).
2) A workstation containing the stored documents such that individual pages could be projected on to a screen at will. A workstation developed by Georges Sebille to store 300,000 pages on twelve 330 mm rolls of film had become quite well-known among documentalists (e.g. Sebille, 1932; Otlet, 1934a). However, the drawing of the Memex in *Life*, drawn by Alfred D. Cimi in consultation with Bush, resembles more closely the workstation of Leonard G. Townsend (1938).
3) Provision for adding new images to the store of microfilm.
4) The ability not merely to locate a known record but to identify and select all and every record with any specified coding.
5) “Associative trails.”

Bush had a low opinion of indexes and classification schemes:

> Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing. When data of any sort are placed in storage, they are filed alphabetically or numerically, and information is found (when it is) by tracing it down from subclass to subclass. It can be in only one place, unless duplicates are used; one has to use rules as to which path will locate it, and the rules are cumbersome. Having found one item, moreover, one has to emerge from the system and re-enter on a new path. (Bush, 1945a, 106).

This account shows no awareness of the ability of indexing and classification systems to bring related material together through collocation and syndetic structure and indicates that Bush's understanding of information retrieval was severely incomplete. Bush declared, in effect, that retrieval should not function as conventional indexes do but as the human brain does, i.e. “as we may think.” Bush thought that the creation of arbitrary associations between individual records was the basis of memory, so he wanted a “mem(ory-)ex”, or “Memex instead of index.” The result was a personalized, but superficial and inherently self-defeating design. Instead of indexing documents directly by their contents or characteristics, Bush proposed coding documents by their perceived relevance to some theme, such as the superiority of the short Turkish bow during the crusades. Documents perceived as relevant to the same theme were thereby linked to each other by a common coding, which provided a “trail” through the collection of documents. In effect, however, each trail *is* an indexing concept and the code for each trail is, in effect, an indexing term.

Relevance judgements, however, are notoriously inconsistent, situational, and likely to change as one's knowledge evolves. Since Bush's trails were based, not on the contents of the document but on perceived relevance of documents to trails, any individual's pattern of trails would be unlikely to be satisfactory for any other user. A personalized information system may be advantageous for an individual but has limited usefulness for others. Furthermore, perceptions of relevance are less stable than perceptions of subject content. For this reason conventional subject-based indexing remains tolerable, indeed preferable, in practice, because a system based directly on relevance, as in the Memex, could not change itself automatically to reflect continuously changing perceptions of relevance. The trails, being based on one individual's personal knowledge and perceptions of relevance, would be highly obsolescent. As a user's knowledge increased, perceptions of relevance would change, and the trails would need to be remade. Any given pattern of trails would remain appropriate only so long as the user did not learn anything from the use of the Memex--or in any other way.

The Technology of the Memex

Bush's Memex draws on two main sources: His view of associative trails as the mechanism by which the brain works; and photographic and other technology available in the late 1930s. The features noted above, and other refinements such as the small camera strapped to the researcher's forehead to photograph anything he or she looked at, were more or less feasible individually with the technology of 1939, although combining them into a single workstation would probably not have been practical.

In 1939 Bush was, in fact, responsible for the design and construction of a document retrieval machine much faster and technologically more advanced than anything previously attempted. This prototype formed the context, the recognizable technological basis (recognizable in Cimi's drawings), and, presumably, the stimulus for his think-piece “As We May Think.”

Microfilm selectors

In the 1920s and 1930s microfilm had become popular as a storage medium for records, especially in banks, and all kinds of people were busily inventing microfilm equipment (Stewart & Hickey, 1960). Microfilming saved storage space and banks found that microfilming cancelled checks was a useful measure against fraud (Johnson, 1932; Schwegmann, 1940). But, since the documents were unlikely to have been microfilmed in an order that was convenient for identifying individual records, the question became how to search for any given document. Punched cards were well known, but the card readers were relatively slow--150 cards per minute--and finding the right card only gave one the address of the document on microfilm. It would be more convenient to have an “integral” retrieval system, one which combined the index and the document. There were two logical possibilities. One could attach frames of microfilm to the card (“aperture cards”) or one could record the logical equivalent of a punched card on to the microfilm alongside the image of the document. One might punch holes in the film or arrange opaque and translucent spots on the film to denote hole or no-hole. Each of these techniques was tried. The usual form of microfilm selector technology is to create a “search card” (a punched card) or template, bearing the coding pattern sought, and align it and the coded areas on the microfilm between a light source and a photoelectric cell. See Figures 1 and 2. As the microfilm bearing codes moves past the search card, the coincidence of a pattern on the microfilm matching the pattern on the search card would affect the flow of light from the light source to the photocell and, thereby, the flow of electric current from the photocell. In this way the desired record is identified and appropriate action, such as the creation of a copy, is triggered.

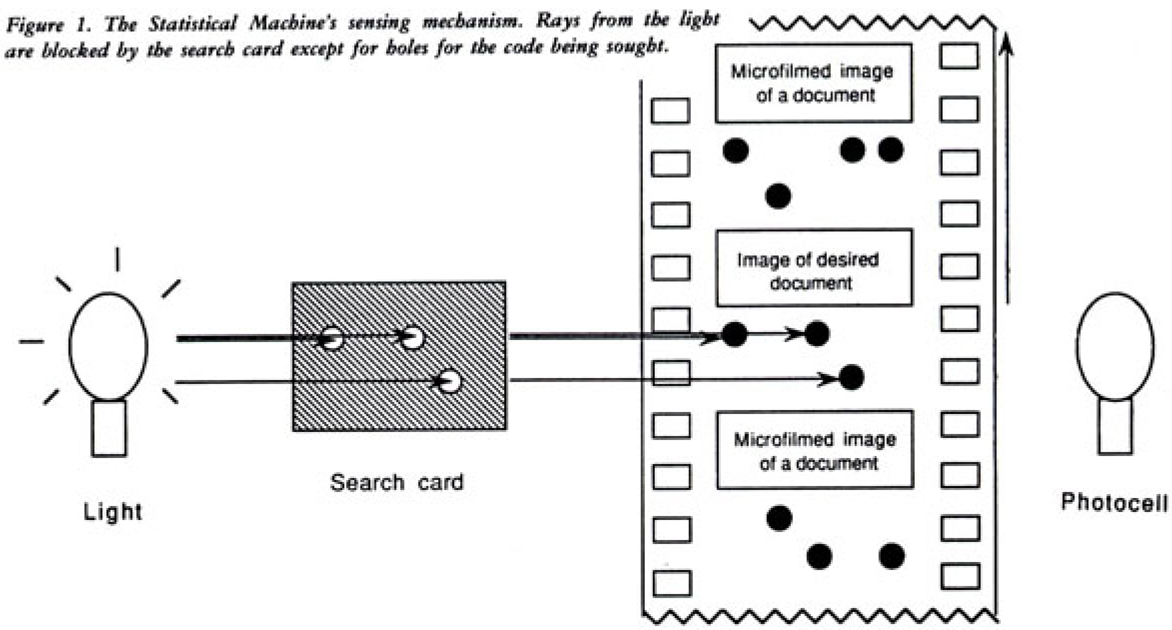

FIG. 1. Microfilm selector sensing mechanism. Rays from the light are blocked by the search card except for holes for the codes being sought. Only when opaque codes on the moving film correspond to the holes on the search card is all light momentarily blocked from the photocell, indicating a hit.

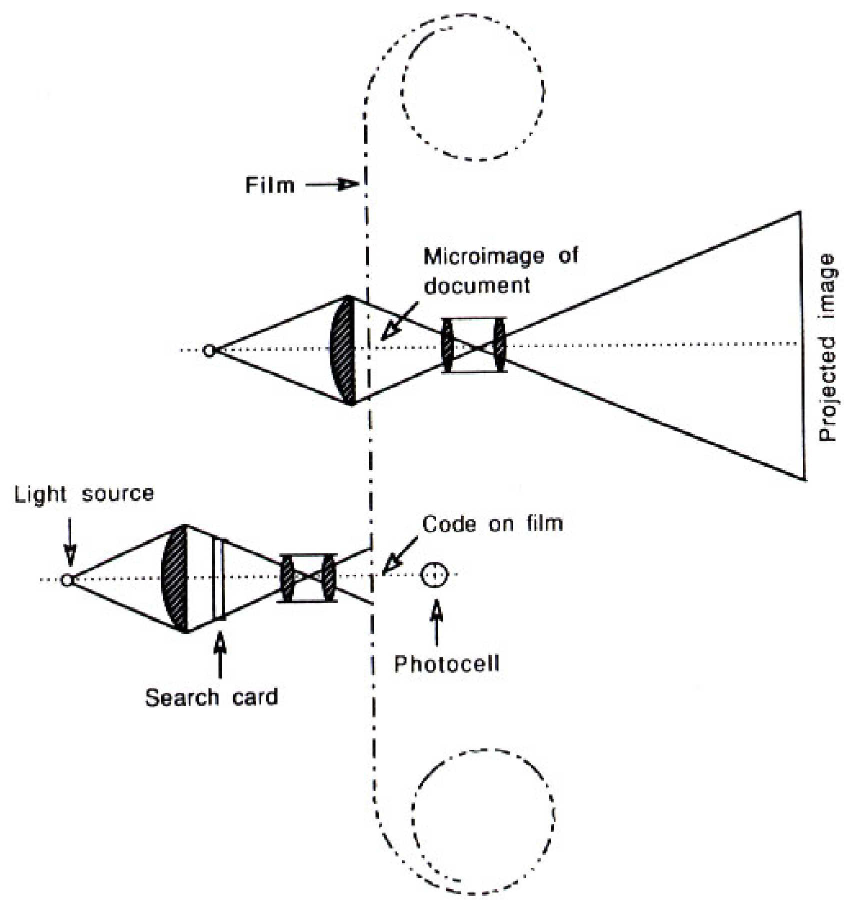

FIG. 2. Diagram of microfilm selector showing sensing mechanism and, above it, projection of copy of the document. [Based on Goldberg, 1932b.]

It is not widely recognized how powerful these machines promised to be. At the 1925 International Congress of Photography a grainless microfilm was demonstrated capable of resolution equivalent to storing the entire text of the Bible fifty times over on one square inch of film. G. W. W. Stevens (1968, 302) gives one million bits per square millimeter of film as the theoretical but, in practice, unachievable limit for microfilm. By 1961 a search speed of 15,000 codes per minute had been achieved on an experimental microfilm selector (Bagg & Stevens, 1962, App. B). Microfilm selectors were developed for small rectangular pieces ("chips") of film as well as for rolls of microfilm.

Giant Brains; or, Machines that Think, a popular introduction to digital computers, assumed, in 1949, that in the future "automated library" catalog records (and, eventually, the documents) would be on microfilm and retrieved by a digital computer:

> You will be able to dial into the catalogue machine `making biscuits.' There will be a flutter of movie film in the machine. Soon it will stop, and, in front of you on the screen will be projected the part of the catalogue which shows the names of three or four books containing recipes for biscuits. (Berkeley, 1949, 181).

Eventually the indexing was moved off the film on to digital computers connected to microfilm readers: Computer Assisted Retrieval (CAR) replaced photoelectric microfilm selectors.

In the literature of librarianship it is sometimes stated that Ralph R. Shaw (1907-1972), a distinguished librarian and professor, invented or "constructed" the microfilm rapid selector (e.g. N. D. Stevens, 1978). But this is inaccurate. The machine publicized by Shaw in 1949 was based on the earlier prototype developed from 1938 to 1940 by a team at Massachusetts Institute of Technology (MIT) under Bush's direction. The project manager for the Bush prototype was John H. Howard and the research assistants were Russell C. Coile, John Coombs, Claude Shannon, and Lawrence Steinhardt. Eastman Kodak and National Cash Register each provided $10,000 funding. The project's objective was to develop, within two years, a prototype machine capable of selecting microfilmed business records from microfilm *rapidly*: A microfilm *rapid* selector. Bush's selector was indeed rapid because it took advantage of two new developments: Improved photoelectric cell technology; and the stroboscopic lamp pioneered by his colleague Harold E. Edgerton. By creating a bright flash of light lasting only one-millionth of a second, the stroboscopic lamp made it possible to copy a selected microfilm image "on the fly," without stopping the film (and the search) to make a copy. The Bush microfilm selector was never used operationally, except that it seems to have been used for cryptanalysis: It was, after all, designed to be effective at identifying (selecting) every occurrence of a specified code.

After World War II, Coombs, Howard, and Steinhardt worked together at Engineering Research Associates (ERA) in St Paul, Minnesota (Tomasch, 1980). Bush was approached by his former students, now at ERA, for support for further work on microfilm selectors. Eventually the U.S. Department of Commerce issued a contract to ERA to build a new microfilm selector. A librarian, Ralph Shaw, then Director of the National Agriculture Library, was funded to encode test material and to test the new ERA machine. Carroll Wilson, who had handled Bush's patents at MIT and was now at the Atomic Energy Commission (AEC), is said to have arranged for the AEC to provide funding to the Department of Commerce for microfilm selector development (Coile, 1990; Engineering Research Associates, 1949; Pike & Bagg, 1962).

There was some criticism of the test and some speculation that both Bush and Shaw doubted the future of microfilm selector technology. Shaw publicized the ERA microfilm rapid selector, making it and himself widely-known internationally (Shaw, 1949a; 1949b). Afterwards he dropped it as unsatisfactory.

There is general agreement in the literature that Bush invented the first microfilm rapid selector. But did he? In 1960 *Fortune* magazine reported that in 1949:

> Not long after a public demonstration of the Rapid Selector, Shaw was visited by an engineer named Goldberg, who had worked before the war with the famous German optical firm of Zeiss-Ikon. Goldberg said he was delighted to see that someone had finally found use for his idea of a

> microfilm record combined with an index code--essentially the same idea as Bush's. Goldberg had received a U.S. patent for the idea in 1931. (Bello, 1960, 167).

Shaw told Goldberg that he had been unaware of Goldberg's work and subsequently mentioned Goldberg's patent (E. Goldberg, 1931) in his two principal papers on the ERA microfilm selector (Shaw, 1949a; 1949b). Later Robert Fairthorne (1958) discussed "As We May Think" in the *Computer Journal*. Fairthorne was critical of Bush's ideas, commented that "few of his suggestions were original," and also mentioned Goldberg's prior work. Fairthorne's paper was reprinted in his book *Towards information retrieval* (1961, 135-146). In 1960, Hawkins, writing in a book edited by Shaw, states:

> Shaw credited Dr. E. Goldberg with the first practical application of electronics to the selection of data on film and Dr. Vannevar Bush with the basic principles of organization of knowledge and the basic electronic system used in the Rapid Selector." (Hawkins, 1960, 145).

Hawkins cites and may only have been summarizing a 1949 paper by Shaw (Shaw, 1949a). The detailed history of microfilm selectors by Bagg and Stevens (1962) mentioned Goldberg's U.S. patent. So there were a few passing references to prior invention by Goldberg available fairly prominently in the information retrieval literature by 1962, but, it seems, almost nothing since. In 1957, in an Israeli technical journal, Neumann complained that Vannevar Bush's account of the Rapid Selector in *Life* did not mention Goldberg "the true inventor, who had actually built and demonstrated such a machine years before." (Neumann, 1957, v).

Several questions arise: Who was Goldberg? What precisely did he invent? Did he actually build a microfilm selector? What, if anything, did Goldberg publish on the subject? Was Bush aware of Goldberg's work? If Bush's microfilm selector inspired his Memex fantasy, did Goldberg's work inspire any comparable visions? Has nobody had enough curiosity to find out? Can we draw any conclusions concerning the sociology or scholarship of information science from the striking contrast between the constant acclaim of Bush and the continuing comparative oblivion of Goldberg?

Emanuel Goldberg

Emanuel Goldberg was born in Moscow, Russia, in 1881. He graduated from Moscow University, moved to Germany, and received his doctorate from the University of Leipzig in 1906 with a dissertation on the kinetics of photochemical reactions. In 1917 he moved to Dresden to become a Director of Ica AG., which was later merged with other firms (Contessa-Nettel, Ernemann, Goertz, and Hahn) to form Zeiss Ikon AG., controlled by the Carl Zeiss Foundation of Jena. Goldberg became the first Managing Director of Zeiss Ikon. Based in Dresden, he was also a Professor in the Institute for Scientific Photography at the Technical University there.

Starting as an undergraduate, Goldberg had already by 1931 contributed a remarkable and brilliant range of insights and inventions, mostly connected with photography. These included a zinc plating process, the first hand-held movie camera (the Kinamo), early involvement in television technology, and a well-received book *Der Aufbau des photographischen Bildes* (translation: *The construction of the photographic image*) (E. Goldberg, 1922). It was also Goldberg who had developed the very high resolution microfilm mentioned above and the technology underlying the microdots later used in espionage (E. Goldberg, 1926; White, 1989; 1990).

1931 International Congress of Photography

The 8th International Congress of Photography, held in Dresden in 1931, must be regarded as a peak in Goldberg's career. It was the proposal presented on behalf of the Committee for Sensitometry by Goldberg and his former instructor Robert Luther for a standard measure of film speeds that became the principal topic of discussion. This proposal led to the adoption of the familiar DIN and ASA film speed ratings.

The congress proceedings were heavily technical. However, a special session of a more popular nature was scheduled for the general public at which Goldberg dazzled the audience. "Dr E. Goldberg ... gave an extremely well illustrated popular lecture on `Fundamentals of Talking Films'" reported the *Journal of the Society of Motion Picture Engineers* (Sheppard, 1932). The *Zeitschrift für angewandte Chemie* commented on Prof. Goldberg's "amazingly simple experiments" (Anon., 1931). In his lecture he demonstrated how the vibrations from the needle playing a phonograph recording of the Egmont Overture could be converted into an electric current which produced patterns on an oscillograph and also fluctuations in a rotating glow-lamp which were then converted, by means of a photoelectric cell, from light back to electric current and fed into a loudspeaker to produce the music recorded on the phonograph record (E. Goldberg, 1932a).

At the congress banquet in the city hall, the prestigious Peligot medal of the French Society for Photography and Cinematography was bestowed on Goldberg. These events seem to have overshadowed a paper that Goldberg presented at one of the technical sessions entitled "Neue Wege der photographischen Registertechnik" ("New Methods of Photographic Indexing"). The paper appeared in the published congress proceedings under the title "Das Registrierproblem in der Photographie" ("The retrieval problem in photography". Goldberg, 1932c). This clear and concise paper describes the design of a microfilm selector using a photoelectric cell. It is, perhaps, the first paper on electronic document retrieval and describes what seems to have been the first functioning document retrieval system using electronics. A prototype was demonstrated. The *British Journal of Photography* recognized the significance of this paper and took the extraordinary step of republishing it in English, with the title literally and unhelpfully translated as "Method of Photographic Registration" (E. Goldberg, 1932b). (See elsewhere in this issue for a new translation).

Two moves that were then unusual in German industry indicate that Goldberg believed that his photoelectric microfilm selector was potentially important: The German patent was issued in his name as well as that Zeiss Ikon and, in renewing his employment contract, he negotiated for a share in any future royalties for this particular patent (H. Goldberg, 1990). Two different prototypes were built (H. Goldberg, 1990), but they were not developed into a product and we have, as yet, found no other record of them. A detailed official history of the achievements and products of Zeiss Ikon and its antecedent firms, issued in 1937, makes no mention of a photoelectric microfilm selector. There is an intriguing reference to "a so-called check and document retrieval machine created for banks and industrial firms which made it possible ... to supply photographic copies" (Zeiss Ikon, 1937, 122. Transl. MKB), but we take this to be a reference to non-electronic equipment, including the Antwerp giro machine noted below.

In 1933, with Hitler in power, Goldberg, being of Jewish origin, was physically assaulted by Nazis, dismissed from his professorship, and became a refugee. He emigrated to Paris, where he directed Optica and Iconta, Zeiss subsidiaries in France from 1933 to 1937. The 1937 official history of Zeiss Ikon, published in the Nazi era, does not mention Goldberg in spite of the leading role he had played for most of the firm's existence. In 1937 he moved on to Tel Aviv, founded Goldberg Instruments Ltd, assisted the Allied war effort, remained interested in information retrieval, and died in 1970.

It was not only the Nazis who obscured Goldberg's achievements. In 1946, twenty years after the publication Goldberg's classic paper on microdot technology (E. Goldberg, 1926), the *Reader's Digest* published a boastful article about espionage and microdots signed by J. Edgar Hoover, Director of the U.S. Federal Bureau of Investigation. Without mentioning the name Goldberg, Hoover writes of "the famous Professor Zapp, inventor of the micro-dot process, at the Technical High School in Dresden." Hoover apparently conflates Goldberg's achievement and academic appointment with the name of Walter Zapp, inventor of the Minox subminiature camera, also used by spies but of a quite different technology and incapable of producing microdots. An erroneous description of microdot production were made follows (Hoover, 1946, 3. For discussion of "this concoction of semitruths and overt disinformation" see White, 1990, 191-195. *Reader's Digest* printed Hoover's piece as the lead paper of the April 1 issue.)

Who Knew What When?

As noted above, Goldberg personally drew Shaw's attention to his patent which was subsequently mentioned by Shaw, Fairthorne, Bagg and Stevens, and, occasionally, in surveys of documentation techniques (e.g. International Federation for Documentation, 1964, 298).

The relevance of Goldberg's patent to document retrieval is not obvious from its title or from the abstract in the U.S. *Patent Gazette*, where it appeared on December 29, 1931. He, like others before Bush, called his invention a "statistical machine." The German patent, submitted in 1927, was entitled "Apparatus for Selecting Statistical and Accounting Data" (Zeiss Ikon & Goldberg, 1938). *Electronics* (1932) included Goldberg's U.S. patent in its regular patent listings with the following summary: "Statistical machine. Use of light beam and phototube for adding, sorting and other statistical operations."

It was also rather unhelpful that Goldberg used the German word "Register" and its derivatives which have multiple meanings to do with recording, indexing, and, in the context of photography and document production, aligning. Worse, the literal 1932 English translation, "Method of Photographic Registration," is quite misleading and does not suggest a paper on document retrieval.

Goldberg's microfilm selector was known at the Kodak Research Laboratories in Rochester, New York, before Kodak funded Bush to construct a "rapid" microfilm selector. Two leading scientists from the Kodak Research Laboratories, Samuel Edward Sheppard and Adrian Peter Herman Trivelli, attended the 1931 Congress, were personal friends of Goldberg, and would presumably have seen the prototype demonstrated. In 1937, Goldberg's son, Herbert Goldberg, went to work at the Eastman Kodak Research Laboratories. In 1938, an employee in the Laboratories, Richard S. Morse, applied for and later assigned to Eastman Kodak a patent for a refinement of the code sensing on what is clearly a microfilm selector (Morse, 1942).

Goldberg's microfilm selector was also known at IBM. James Ware Bryce (1880-1949), the Chief Scientific Director of IBM during the 1930s, monitored new developments in electronics and was interested in microfilm as a data storage medium. (For Bryce see Anon., 1949, and Bashe et al., 1986). When Goldberg's U.S. patent appeared in 1931, IBM promptly acquired a license for it. One of Bryce's own 400 patents, applied for in 1936, was for an advanced microfilm selector (Bryce, 1938).

Knowing of Goldberg's patent, one can find the congress paper, if one would recognize it from its title, by looking under Goldberg in the *Internationale Bibliographie der Zeitschriftenliteratur* (*IBZ* or "*Dietrich*"), (Abt. A, v. 70, 1932), the principal German index to periodical literature, or in the entry for Goldberg in the 1937 edition (but not the 1970 edition) of a leading directory of scientists, *J. C. Poggendorff* (1937). It would not have been difficult, for example, for Shaw, who was fluent in German and an expert on bibliography, to have found it. Nevertheless this paper appears to have remained almost totally unknown and uncited.

In the U.S.A. the *Journal of Documentary Reproduction*, published from 1938 to 1942, reported research and development related to microforms in documentation. An extensive bibliography in the first issue includes both the English and the German versions of Goldberg's congress paper and another unhelpful literal translation of the German version: "Problem of Registration in Photography" (Berthold, 1938, 100). No other reference to Goldberg has been noticed in the *Journal*, although there is a statement, without sources or explanation, that "Work has already been done on selecting devices of the stroboscopic and photocell types" (Carruthers, 1938, 269).

In 1937 a report on entitled *The Present State of Equipment and Supplies for Microphotography* was prepared by V. D. Tate for the Committee on Scientific Aids to Learning of the National Research Council. Bush was a member of this committee. The report, reprinted as a special issue of the *Journal of Documentary Reproduction* (Tate, 1938), includes a brief description of work by Merle E. Gould (given as Merle C. Gould):

> Photo-electric cells, in conjunction with a keying system incorporated on the original microfilm, would permit machine selection of predetermined types... The preliminary models showed considerable promise (Tate, 1938, p. 48)

One of Gould's patents, applied for in 1936, was for an “Identifying means,” using an array of photoelectric cells to detect specified patterns of light, the same general approach as in the subsequent Bush microfilm rapid selector (Gould, 1940). Bush, as a member of the committee, could be expected to have read the report and, therefore, to know of Gould's prior work. Goldberg is not mentioned in the report.

Soon after the War a microfilm selector development project at ERA was given the code name “Goldberg,” which might perhaps have been a reference to Emanuel Goldberg (Burke, 1991). If this code name had some other referent, such as Rube Goldberg, it would have been an ironic coincidence.

The European Documentalists

How well was Goldberg's machine known to European specialists in information retrieval (then called “documentation”) in the 1930s? If Bush's prototype microfilm selector inspired his Memex vision, did Goldberg's earlier prototype also inspire any similar, earlier visions of personal information machines among the documentalists? The literature on documentation in the 1930's was as preoccupied with microfilm technology as it is now with computer technology and for the same reason, each being the most promising information retrieval technology of the time. The principal international journal, published by the International Institute for Documentation (I.I.D., now the International Federation for Information and Documentation (F.I.D.)), was the *I.I.D. Communicationes*, in which two related inventions from Goldberg's firm received attention. One was a microfilm reader built for the Antwerp Giro administration, which enabled the visual locating and scanning of a microfilm of pre-sorted cancelled checks at up to 3,000 checks per minute (Keegstra, 1933). A photo of the machine can be seen in the *I.I.D. Communicationes* 1, Fasc. 3 (1934): Plate XLV. Another was a microfilm application that reduced labor costs and clerical errors in preparing telephone subscribers' monthly invoices at the Amsterdam city telephone system (Maitland, 1931). A picture of the equipment also appears in *I.I.D. Communicationes* 1, Fasc. 3 (1934): Plate XLIV. Goldberg's prototype microfilm selector would certainly have been of considerable interest had it been known, but there seems to be little evidence that it was.

The most thorough treatment of information retrieval in that period was Paul Otlet's *Traité de documentation* (1934). Otlet's idiosyncratic text is quite forward-looking. In television he recognizes the potential for using telecommunications for remote access to documents:

> Soon television will be a problem that has been essentially solved, as it already is in theory: the image is reproduced at a distance without a wire. One can imagine the electric telescope, permitting one to read at home, “tele-reading” the books set out in the reading rooms of the great libraries, at the pages requested in advance. (Otlet, 1934, p. 238, transl.).

Later he enumerates inventions, such as machine translation, that are needed for information retrieval and information processing. After stressing the importance of telecommunications and the need for technical standards, Otlet provides a concise outline of a personal information system, including an anticipation of hypertext:

> We should have a complex of associated machines which would achieve the following operations simultaneously or sequentially: 1. Conversion of sound into text; 2. Copying that text as many times as is useful; 3. Setting up documents in such a way that each datum has its own identity and its relationships with all the others in the group and to which it can be re-united as needed; 4. Assignment of a classification code to each datum; [division of the document into parts, one for each datum, and] rearrangement of the parts of the document to correspond with the classification codes; 5. Automatic classification and storage of these documents; 6. Automatic retrieval of these documents for consultation and for delivery either for inspection or to a machine for making additional notes; 7. Mechanized manipulation at will of all the recorded data in order to derive new combinations of facts, new relationships between ideas, new operations using symbols.

> The machinery which would achieve these seven requirements would be a veritable mechanical and collective brain. (Otlet, 1934, p. 391, transl.)

Otlet, who had himself been a pioneer in the use of microforms for documentation and had been active in earlier International Congresses on Photography, probably knew of Goldberg and of Goldberg's high resolution microfilm. However Otlet's brief discussion of "selecting machines" (p. 390) does not extend beyond punched card equipment. There seems to be no recognizable allusion to photoelectric retrieval from microfilm in the *Traité*.

The German version of Goldberg's congress paper is included, without annotation, in an anonymous bibliography in the *I.I.D. Communicationes* in 1935 (Anon., 1935, 19). It also occurs in a bibliography compiled by Walter Schürmeyer and T. P. Loosjes (1937) also published in the *I.I.D. Communicationes* in 1937. But Goldberg is not mentioned in Loosjes' subsequent textbook on documentation even though it includes historical background and does discuss microfilm selectors (Loosje, 1967), which suggests that it was Schürmeyer who had added the Goldberg paper to the bibliography.

Walter Schürmeyer was librarian of the Kunst und Technik Bibliothek in Frankfurt, Germany, during the 1930s and was active in documentation circles (Habermann, Klemmt, & Siefkes, 1985, 315-316). He chaired an I.I.D. committee on technical methods in documentation. In a paper presented to the 29th conference of the German Librarians Association (VDB) in Darmstadt in 1933, and published in the *Zentralblatt für Bibliothekswesen*, a internationally a leading journal of librarianship, Schürmeyer (1933) drew attention to the Zeiss Ikon Giro check machine as suitable for library catalogs and to the Contax 35 mm camera developed at Zeiss Ikon under Goldberg's supervision. At the 1935 International Congress on Documentation in Copenhagen, Schürmeyer (1936) presented a paper on the use of microfilm which contains a paragraph on photoelectric selection techniques which contains no citations but which seems to refer to Goldberg's selector:

> Such film-based documentation can be completely automated using photoelectric selection methods. This new discovery permits any signals in the form of dots, and also numerals or letters, recorded on transparent sheets or films to be scanned and selected with the help of a photoelectric cell. It is to be expected that we shall soon have an automated machine which, through switching a photoelectric cell to a given code, stops stored films with the same code, makes copies and stores them again in the same sequence. Thus, in a very short time a bibliography, a summary of available references and other documents including illustrations could be compiled completely automatically." (Col. Schü. 8, transl.)

Schürmeyer also foresaw the potential of telecommunications:

> But what a revolution for information retrieval and especially for libraries television can bring! Perhaps one day we shall see our reading rooms deserted and in their place a room without people in which books requested by telephone are displayed, which the users read in their homes using television equipment." (Col. Schü. 9, transl.).

Otlet, the bibliographer, and Schürmeyer, the librarian, were more forward-looking in their ideas about information retrieval technology than was Bush, the professor of electrical engineering, a decade later.

Bush Reassessed

Bush seems to have said little in his published work about the antecedents of his Memex or of his microfilm rapid selector. Three considerations suggest that he was unaware of the detail of Goldberg's work when he built his prototype in 1938-40:

1) Russell C. Coile, a research assistant in the development of Bush's microfilm rapid selector, does not recall any reference to Goldberg's work at the time (Coile, 1990).

2) Shaw, whose involvement was a direct development of Bush's work, told Goldberg in 1949 that he had been unaware of Goldberg's work (Bello, 1960; H. Goldberg, 1990).
3) A difference in technical design. Goldberg matched by complement or “extinction”: By matching translucent codes on the search card against opaque marks on the film (as in Fig. 1), the occurrence of the sought-for code would be detected by the momentary blocking of all light to the single photocell. Bush's design, however, matched by coincidence: Holes as codes on the search card were matched against translucent codes on the film. Occurrence of the sought-for code was detected by sensing the presence of light in each and every position prescribed by the search template but in no other positions. A bank of photoelectric cells, one for each possible position, was needed, and just the right combination of electric outputs had to be detected. This was a much more difficult task and a less elegant design than Goldberg's. Subsequent designers searched by complement (extinction) as, we speculate, Bush would have done had he known the details of Goldberg's work.

Nevertheless, Goldberg's rapid selector was known in at least two leading U.S. research centers: at IBM and at Eastman Kodak Research Laboratories, one of Bush's funders. It should also be remembered that new developments are often known quite widely in vague and incomplete ways. We speculate that Bush did not independently originate the notion of an electronic microfilm selector, although that was possible. It is not surprising that the same invention sometimes occurs independently and more or less simultaneously when a need is present and the technology becomes ripe. Inventors prefer inventing to copying. With the benefit of hindsight, the invention of the photoelectric microfilm selector seems almost inevitable since it was a logical development of each of two different technologies: (i) It was a logical extension of punched cards to transpose the coding on to the microfilm, then the one outstanding medium for compact storage; and (ii) it was very similar to sound movies where a coding to be scanned is placed on the film alongside the images on the film. In such circumstances, multiple independent invention is to be expected. For example, it appears that Helen M. Davis, working with her husband, Watson Davis, and Rupert H. Draeger, invented a microfilm selector independently in 1935 (Bagg & Stevens, 1962, 17). Watson Davis, like Schürmeyer, presented a paper at the 1935 Copenhagen I.I.D. conference in which he also referred, rather vaguely, to photoelectric selection techniques and retrieval from microfilm.

Watson Davis occupied an interesting position in all this. He was very interested in the use of microfilm for document storage and dissemination. He knew Bush, Goldberg, Otlet, Schürmeyer and C. E. K. Mees, head the Eastman Kodak Research Laboratories, personally. He probably knew, or knew of, most, if not all, of the other individuals mentioned in this paper. (For Davis see Farkas-Conn, 1990).

Conclusions

Our inquiry into the background and antecedents of Vannevar Bush's Memex and of his microfilm rapid selector is certainly incomplete. However, the evidence found indicates that the generally accepted view needs substantial revision.

Emanuel Goldberg designed a photoelectric microfilm selector, which he called a “statistical machine”, by May 1927. Two prototypes were built at Zeiss Ikon by 1931 and, perhaps, constitute the first successful electronic document retrieval. Microfilm selector technology was known in at least two leading research centers in the U.S.A. (Eastman Kodak and IBM) by 1931 or shortly thereafter and in both cases a direct connection to Goldberg can be shown. This technology was reported at international congresses in 1931 and 1935 and a number of U.S. inventors were working on it by 1938 (e.g. Bryce, H. Davis, Gould, and Morse).

Vannevar Bush's contributions in this area were two-fold: (i) A significant engineering achievement by the team under his leadership in building a truly *rapid* prototype microfilm selector; and (ii) a speculative article, “As We May Think,” which, through its skillful writing and the social prestige of the author, has had an immediate and lasting effect in stimulating others. As Fairthorne observed, Bush's paper was timely and "opened people's eyes and purses."

The prewar information retrieval specialists of continental Europe, the "documentalists," largely disregarded by post-war information retrieval specialists, had ideas that were considerably more advanced than is now generally realized.

Acknowledgements

The author gratefully acknowledges the assistance of Thomas C. Bagg, Samia Benidir, Colin Burke, Russell C. Coile, Horace P. Flatt, Herbert Goldberg, Mead Kibbey, Calvin N. Mooers, W. Boyd Rayward, William White, and Wolfgang Welkerling, and would welcome further information and suggestions.

Appendix

Sources

Information on Goldberg is not easily found. The best sources identified were a short appreciation by Neumann (1957) in a Goldberg Festschrift issue of *Bulletin of the Research Council of Israel*, an 1969 interview reported in *Popular Photography* (N. Goldberg, 1969), and an entry in the *International Biographical Dictionary of Central European Emigrés* (1983, v. 2, 388). Other short biographical sources include the vita printed at the end of his doctoral dissertation (E. Goldberg, 1906, 46), Browne and Partnow (1983, 234-235), Gubas (1985a; 1985b), Kaprelian (1971), *J. C. Poggendorff* (1937; 1970), Sipley (1965, 58-59), *Wer ist's?* (1928, ?), and *Zeitschrift für wissenschaftliche Photographie* (Anon., 1957). Several of his publications can be found by combing through contemporary issues the *IBZ* ("*Dietrich*"). Aspects of his work can be found through the indexes of standard scientific works on or including photography, e.g. Eder (1945), Mees (1966), and White (1990).

Shaw (1949a) provides a convenient introduction to microfilm selectors. A postscript to this paper by E. M. R. Ditmas incorrectly cites the technical report by Engineering Research Associates (1949) on the ERA microfilm rapid selector as PB 97 535 instead of PB 97 313, an error repeated by some subsequent writers. Bagg and Stevens (1961) provide the best historical account of microfilm selector development, albeit incomplete with respect to Goldberg, which can be supplemented by the later account by Alexander and Rose (1964). G. W. W. Stevens (1968, chap. 12) provides a summary, as does the International Federation for Documentation (1964, chap. 9).

References

Alexander, S. N., & Rose, F. C. (1964). The current status of graphic storage techniques: Their potential application to library mechanization. In *Libraries and automation*, ed. by B. E. Markuson. (pp. 111-40). Washington, D.C.: Library of Congress.

Anon. (1931). Versammlungsberichte. *Zeitschrift für angwandte Chemie*, *44*: 875-78.

Anon. (1935). Ergänzungen und Verbesserungen zu dem Artikel von Dipl. Ing. J. P. C. van Asperen über moderne photgraphische Reproduktionsverfahren. *I. I. D. Communicationes*, *2*, Fasc. 2: 16-19.

Anon. (1945). A machine that thinks. *Time 46*:93-94.

Anon. (1949). The light he leaves behind. *Think* April 1949: 5-6, 30-31.

Anon. (1957). Professor Dr. Goldberg 75 Jahre. *Zeitschrift für wissenschaftliche Photographie*, *52*: 105-106.

Bagg, T. C., & Stevens, M. E. (1961). *Information selection systems retrieving replica copies: A state-of-the-art report*. (National Bureau of Standards Technical note 157). Washington, D.C.: Government Printing Office.

Bashe, C. J. *et al*. (1986). *IBM's early computers*. Cambridge, MA: MIT Press.

Bello, F. (1960). How to cope with information. F*ortune* (1960, September): 162-67, 180, 182, 187, 189, 192.

Berkeley, E. C. (1949). *Giant Brains: or, Machines that think*. New York: Wiley.

Berthold, A. (1938). Selected biography on photographic methods of documentary reproduction. *Journal of Documentary Reproduction*, *1*, 87-123.

Browne, T., & Partnow, E. (1983). *Macmillan biographical encyclopedia of photographic artists & innovators*. New York: Macmillan.

Bryce, J. W. (1938). *Statistical machine*. U.S. Patent 2,124,906. July 26, 1938.

Burke, C. (1991). Personal communication.

Bush, V. (1945a). "As we may think." *Atlantic monthly 176*, 101-108.

Bush, V. (1945b). As we may think: A top US scientist foresees a possible future world in which man-made machines will start to think. *Life 19*(11), 112-114, 116, 118, 123-124.

Bush, V. (1946). *Endless horizons*. Washington, D.C.: Public Affairs Press.

Bush, V. (1967). *Science is Not Enough*. New York: Morrow.

Bush, V. (1970). *Pieces of the Action*. New York: Morrow.

Carruthers, R. H. (1938). The place of microfilm in public library reference work. *Journal of Documentary Reproduction*, *1*, 263-268.

Coile, R. C. (1990). Personal communication.

Eder, J. M. (1945). *History of photography*. New York: Columbia University Press.

Electronics (1932) *4*(1), 35.

Engineering Research Associates, Inc. (1949). *Report for the microfilm rapid selector*. St. Paul, MN: Engineering Research Associates. (PB 97 313 often misscited as 97 535).

Fairthorne, R. A. (1958). Automatic retrieval of recorded information. *Computer Journal*, *1*, 36-41.

Fairthorne, R. A. (1961). *Towards information retrieval*. London: Butterworths.

Farkas-Conn, I. (1990). *From documentation to information science*. New York: Greenwood.

Goldberg, E. (1906). *Beiträge zur kinetik photochemischer Reaktionen: Inaugural-Dissertation ... Universität Leipzig*. Leipzig: Barth.

Goldberg, E. (1922). *Der Aufbau des photographischen Bildes*. Halle: Knapp. 2nd ed., 1925.

Goldberg, E. (1926). A new process of micro-photography. *British Journal of Photography*, *73*, 462-465.

Goldberg, E. (1931). *Statistical Machine*. U.S. patent 1,838,389. Dec. 29, 1931.

Goldberg, E. (1932a). Die Grundlagen des Tonfilms. In The 8th International Congress of Photography (Dresden, 1931). *Bericht über den VIII. internationalen Kongress für wissenschaftliche und angewandte Photographie*. Herausg. von J. Eggert und A. v. Biehler. (pp. 213-214). Leipzig: Barth, 1932.

Goldberg, E. (1932b). Methods of photographic registration. *British Journal of Photography*, *79*, 533-534.

Goldberg, E. (1932c). Das Registrierproblem in der Photographie. In The 8th International Congress of Photography. (Dresden, 1931). *Bericht über den VIII. internationalen Kongress für wissenschaftliche und angewandte Photographie*. Herausg. von J. Eggert und A. v. Biehler. (pp. 317-320). Leipzig: Barth, 1932.

Goldberg, H. (1990). Personal communication.

Goldberg, N. (1969). The other Goldberg: A visit with Zeiss Ikon's practical prodigy. *Popular Photography*, *65*(5), 88-89, 154.

Goldschmidt, R. B., & Otlet, P. (1925). *La conservation et la diffusion de la pensée: Le livre microphotique*. (I.I.B. publication, 144). Brussels: Institut International de Bibliographie, 1925. English transl. in Otlet, P. (1990). *International organization and dissemination of knowledge: Selected essays*. Transl. by W. B. Rayward (ed.)(pp. 204-210). Amsterdam: Elsevier.

Gould, M. E. (1941). *Identifying means*. U.S. patent 2,231,186. February 11, 1941.

Gubas, L. (1985a). Emmanuel Goldberg. *Zeiss Historica*, *7*.

Gubas, L. (1985b). Who invented the Contax? *Zeiss Historica*, *7*.

Habermann, A., Klemmt, R., & Siefkes, F. (1985). *Lexikon deutscher wissenschaftlicher Bibliothekare 1925-1980*. Zeitschrift für Bibliothekswesen und Bibliographie Sonderheft 42. Frankfurt: Klostermann.

Hawkins, R. R. (1960). *Production of micro-forms*. New Brunswick, NJ: Graduate School of Library Service, Rutgers--The State University. State of the library art, ed. by R. Shaw, vol. 5, part 1.

Hoover, J. E. (1946). The enemy's masterpiece of espionage. *Reader's Digest 48*, 1-6.

International Biographical Dictionary of Central European Emigrés, 1933-1945. (1983). Munich: Saur.

Internationale Bibliographie der Zeitschriftenliteratur. (1932). Abt. A, v. 70. ("Dietrich")

International Federation for Documentation. (1964). *Manuel practique de reproduction documentaire et de sélection*. FID Publ. 353. Paris: Gauthiers-Villars.

J. C. Poggendorff's biographisch-literarisches Handwörterbuch. (1937). Berlin: Verlag Chemie. *VI*, ii, 916; Also (1970) *VIIb*, Teil 3, 1685.

Johnson, W. E. (1932). Protection and profits through photography. *Bankers Magazine*, *125*, 537-540.

Kaprelian, E. K. (1971). In Memoriam: Emmanuel Goldberg, 1881-1970. *Photographic Science and Engineering*, *15*, 3.

Keegstra, H. (1933). Die Photographie in der Gemeinde-Giroverwaltung. In International Institute for Documentation. *XIIe Conférence. Rapports*. (pp. 130-134). (I.I.D. Publ. 172a). Brussels: I.I.D.

Lilley, D. B., & Rice, R. W. (1989). *A history of Information Science 1945-1985*. San Diego, CA: Academic Press.

Loosjes, T. P. (1967). *On documentation of scientific literature*. London: Butterworths.

Maitland, C. E. A. (1931). Die photographische Gesprächszahlerablesung und die optisch-mechanische Auswertung der Zählerstände. *Zeitschrift für Fernmeldetechnik, Werk- und Gerätebau*, *12. Jg*, 2. Heft (28 Feb. 1931): 17-20.

Mees, C. E. K. (1966). *The theory of the photographic process*. (3rd ed.) New York: Macmillan.

Morse, R. S. *Rapid selector-calculator*. U.S. Patent 2,295,000. September 8, 1942.

Neumann, S. (1957). Prof. Emanuel Goldberg. *Bulletin of the Research Council of Israel*. Section C: Technology, *5C*(4), iii-v.

Nyce, J. M., & Kahn, P. (1989). Innovation, pragmaticism, and technological continuity: Vannevar Bush's Memex. *Journal of the American Society for Information Science*, *40*, 214-220.

Otlet, P. (1934a). Le livre photomicrographique. [Followed by] Notice sur les brevets Georges Sebille. *I.I.D. Communicationes 1*(Fasc. 1): 19-23.

Otlet, P. (1934). *Traité de documentation*. Brussels: Editiones mundaneum. Reprinted Liège, Belgium: Centre de lecture publique de la communauté Française, 1989.

Pike, J. L. & Bagg, T. C. (1962). The rapid selector and other NBS document retrieval studies. In V. D. Tate (Ed.), *Proceedings of the Eleventh Annual Meeting* (pp. 213-227). Annapolis, MD: National Microfilm Association.

Rayward, W. B. (1976). *The universe of information: The Work of Paul Otlet for Documentation and International Organisation*. (FID Publ. 520). Moscow: VINITI.

Rayward, W. B. (1986). Otlet, Paul-Marie-Ghislain. In *ALA encyclopedia of library and information services*. 2nd ed. (pp. 626-628). Chicago: American Library Association.

Schuermeyer, W. (1933). Die Photographie im Dienste der bibliothekarischen Arbeit. *Zentralblatt für Bibliothekswesen 50*, 580-583.

Schuermeyer, W. (1936). Mitteilungen über einige technische Neuerungen und Anwendungsmethoden fotographischer Hilfegeräte für das Dokumentarische Arbeiten. *I.I.D. Communicationes*, *3* (fasc. 1), cols. Schü. 1-10.

Schuermeyer, W., & Loosjes, T. P. (1937). Literatur ueber die Anwendung von photographischen Reproduktionsverfathren in der Dokumentation. *I.I.D. Communicationes*, 4 (fasc. 3), 23-29.

Schwegmann, G. A. (1940). Microfilming in business and industry. *Journal of Documentary Reproduction 3*, 147-152.

Sebille, G. (1932). *Method and apparatus for reading books and the like*. U.S. Patent 1,889,575. Nov. 29, 1932.

Shaw, R. R. (1949a). The rapid selector. *Journal of documentation*, *5*, 164-171.

Shaw, R. R. (1949b). Machines and the bibliographical problems of the twentieth century. In L. N. Ridenour, et al. (Eds.), *Bibliography in an age of science* (pp. 37-71). Urbana: University of Illinois Press.

Sheppard, S. E. (1932). Resumé of the Proceedings of the Dresden International Photographic Congress. *Journal of the Society of the Motion Picture Engineers*, *18*, 232-41.

Sipley, L. W. (1965). *Photography's great inventors*. Philadelphia: American Museum of Photography.

Smith, L. C. (1981). 'Memex' as an image of potentiality in information retrieval research and development. In Oddy, R. N. et al. (Eds.). *Information retrieval research* (pp. 345-69). London: Butterworths.

Stevens, G. W. W. (1968). *Microphotography: Photography and photofabrication at extreme resolution*. (2nd ed.) New York: Wiley.

Stevens, N. D. (1978). Shaw, Ralph Robert (1907-1972). In *Dictionary of American library biography*. (pp. 476-81). Littleton, CO: Libraries Unlimited.

Stewart, J. & Hickey, D. (1960). *Reading devices for micro-images*. New Brunswick, NJ: Graduate School of Library Service, Rutgers -- The State University. State of the library art, vol. 5, part 2.

Tate, V. D. (1938). The present state of equipment and supplies for microphotography. *Journal of Documentary Reproduction 2*, 3-62.

Tomasch, E. (1980). The start of an ERA: Engineering Research Associates, Inc., 1946-1955. In N. Metropolis, J. Howlett, & G.-C. Rota (Eds.). *A history of computing in the twentieth century* (pp. 485-495). New York: Academic Press.

Townsend, L. G. (1938). *Method of and apparatus for the indexing and photo-transcription of records*. U.S. Patent 2,121,061. June 21 1938.

Wer ist's? (1928). 9. Ausg. Herausg. von H. A. L. Degener. Berlin: Degener.

White, W. (1989). The microdot: Then and now. *International Journal of Intelligence and Counterintelligence*, *3*, 249-269.

White, W. (1990). *Subminiature photography*. Boston: Focal Press.

Zeiss Ikon AG. (1937). *75 Jahre Photo- und Kinotechnik; Festschrift herausgegeben anlässlich der Feier des 75-jährigen Bestehens der Zeiss Ikon AG. und ihrer Vorgängerfirmen 1862 - 1937*. [?Dresden: Zeiss Ikon].

Zeiss Ikon AG. & Goldberg, E. (1938). *Vorrichtung zum Aussuchen statistischer und Buchhalterischer Angaben*. [German] Patentschrift 670 190. Dec. 22, 1938.