

F011
002
f

18558



OFICINA DE EDUCACION
IBEROAMERICANA

REPÚBLICA ARGENTINA
SECRETARÍA DE ESTADO DE CULTURA Y EDUCACION

UNION PANAMERICANA
DEPARTAMENTO DE ASUNTOS
EDUCATIVOS

CURSO LATINOAMERICANO DE DOCUMENTACION E INFORMACION EDUCATIVA

LA DOCUMENTACION AUTOMATICA
SISTEMAS DE SELECCION

ALAN M. REES

SIE/DOC/CLDIE/DP/4 (200)



centro nacional de documentación e información educativa

INV	018558
SIG	F64
LIB	002 7

De: Curso de Formación Acelerada dictado por la
Asociación Interamericana de Documentalistas y
Técnicos de la Información. París - Francia

"Por mucho tiempo aún, la principal ventaja de los mecanismos automáticos en las bibliotecas se derivará de la aplicación de los principios de su construcción, más que de los mismos autómatas" (R. A. Fairthorne)

"En la actualidad, mucha gente espera el código o la máquina perfectos. En la selección documental, lo importante es lanzarse" (Vocero de una compañía de productos farmacéuticos).

"Nosotros creemos firmemente en el R. I. O. T. (Retrieval in our Time)". Anónimo.

I. INTRODUCCION

La razón que ha llevado a mecanizar la selección de los documentos, se halla evidentemente ligada a la idea universalmente admitida de que se ha producido una crisis en el dominio de la documentación como consecuencia de la extraordinaria expansión de la literatura, hecho que exige el empleo de medios poco tradicionales y tal vez extraordinarios para la organización, mantenimiento y selección de la información. En 1945, Vannevar Bush ya había prevenido a los documentalistas que los métodos para obtener el detalle que nos interesa en un momento dado a través de un montón de documentos, son los mismos que se empleaban desde la época de la navegación a vela. (1). Esta observación fue el punto de partida de un movimiento de descontento hacia las técnicas tradicionales de mantenimiento y de selección de la información. Durante una década, sino más, algunos se regocijaron con las hermosas esperanzas que prometían de antemano las máquinas, en especial las computadoras, y ya se preveía una nueva era en la tecnología de la información. Las tentativas por mecanizar completamente la selección, el análisis, la codificación, el mantenimiento y la búsqueda de documentos no resultaron ni felices ni particularmente logrados. Hoy se reconoce que es preciso planificar cualquier tarea antes de poder utilizar la computadora, por no poder ésta definir las operaciones del espíritu antes que las mismas le sean delegadas. La computadora no puede simular ser una persona humana más que en un sentido muy limitado, puesto que también es limitada la capacidad del hombre para describir lo que hace, de saber por qué lo hace y de enseñar a un mecanismo autónomo cómo hacerlo en su lugar.

No es sorprendente, pues, que se haya llegado tan sólo a automatizar algunos elementos no esenciales de las operaciones relativas a la selección de documentos, su registro alfabético, su clasificación, su codificación, el análisis de temas, la búsqueda y difusión de informaciones. En cambio se han podido efectuar algunas tareas secundarias por vía automática. En la actualidad, hay varios sistemas que permiten la búsqueda, la clasificación, la difusión y el análisis por medios enteramente automáticos.

El presente trabajo tratará, pues, de la mecanización de sistemas secundarios más que de la mecanización en sentido estricto; de los sistemas de mantenimiento y

de búsqueda de las informaciones.

II. SISTEMAS AUTOMATICOS DE SELECCION

La mecanización de la selección de las informaciones es relativamente simple. Una vez que se ha establecido la señalización de los documentos y una vez que se los ha registrado en un medio fácilmente localizable (cinta magnética, tarjetas perforadas, microfichas, etc...) y así como la presentación del pedido de información sea representado de manera conveniente, cualquier máquina de selección de gran velocidad podrá efectuar comparaciones y aproximaciones.

Se podría citar muchos ejemplos de sistemas de operación. La División de Propulsión Aérea de la "General Electric Company" de Evandale, Ohio, se sirve desde 1958 de una computadora para el trabajo de un fichero ordenado. La máquina proporciona una lista de números de documentos, siendo impresos los resúmenes analíticos de los mismos a partir de una cinta magnética separada.

La corrección de tales resúmenes analíticos es verificada antes de ser expedidos a los solicitantes (2).

El Centro de Documentación para la Defensa (antes A. S. T. I. A.) que es el depósito documental para todas las relaciones de búsqueda que realiza por contrato el Servicio de Defensa Nacional y otras numerosas agencias gubernamentales, hace operar un sistema de búsqueda mecanizada con la ayuda de una computadora Univac SS-90. El análisis de los documentos es realizado a mano, con referencia a un Léxico DDC. Los descripciones son coorsionados en la investigación. (3).

La American Society for Metals dirige desde 1960 un Servicio de Documentación que brinda informaciones sobre problemas actuales y realiza investigaciones en la literatura retrospectiva en materia de metalurgia y ciencias auxiliares. El Sistema de Selección fue creado por la Western Reserve University (Center for Documentation and Communication Research). El mismo emplea extractos telegráficos, "role indicators" y factores semánticos. La búsqueda se realiza con una computadora G. E. 225. La máquina saca números y documentos. Los resúmenes analíticos apropiados, conservados en un fichero auxiliar son reproducidos para el solicitante. El fichero completo con alrededor de 125.000 resúmenes analíticos es operado de manera secuencial para cada pedido de búsqueda retrospectiva. La catalogación, la formulación de la investigación, la evaluación de la exactitud de las respuestas son efectuadas a mano (4).

En la Biblioteca Nacional de Medicina se han puesto en práctica otros sistemas de investigación mecanizada. MEDLARS (Medical Literature Analysis and Retrieval System) realizará dentro de poco algunas búsquedas a la vez retrospectivas y de actualidad con ayuda de una computadora Honeywell 800 (5). La Allison Division of General Motors (6) Dupont (7). The American Petroleum Institute (8) y muchas otras organizaciones han creado sistemas de investigación mecanizados.

Es preciso reconocer que en todos esos casos la computadora electrónica no ha contribuido en nada a la mejora de la investigación. La mecanización ha tenido

por resultado el dar vuelta más rápido la página, habiendo cumplido la computadora el papel de depósito automático. La catalogación, el análisis y la formulación de las cuestiones son efectuadas con ayuda de las máquinas. Disponemos de pocos datos sobre la eficacia relativa de la investigación mecanizada y manual.

El encuentro de los pedidos con los documentos depositados plantea algunos problemas que afectan considerablemente la buena operación de los sistemas de búsqueda. Proporcionar los documentos que encuentren exactamente los criterios de investigación pedidos, no es garantía de una perfecta selección de todos los documentos pertinentes. A tal respecto, se ha puesto en duda la eficacia de la lógica de Boole (9). Para superar esa dificultad ha sido propuesta una catalogación probabilística. Los catalogadores atribuyen un peso a cada término empleado en la señalización de un documento. Para una solicitud determinada, los términos asignados en el catálogo a los documentos son pesados y esos pesos empleados para calcular un valor de pertenencia para cada documento. En consecuencia resulta posible obtener una serie de documentos que satisfagan a un pedido, clasificados según su valor de pertenencia (10)

III. SISTEMAS DE CATALOGACION AUTOMATICA. -

La catalogación ha sido considerada como el principal cuello de botella de los sistemas de selección. Por otro lado, se ha demostrado que los métodos empleados para efectuar catalogaciones no permiten en muchos casos que los solicitantes formulen sus preguntas. Las tentativas de mecanizar la catalogación se originan en el deseo de aumentar la eficacia del trabajo sustituyendo los lentos y costosos esfuerzos de los seres humanos por una rutina mecanizada. Tales tentativas se hallan asimismo motivadas por la idea de que la catalogación de las informaciones efectuadas a partir del lenguaje natural evita la intervención del juicio subjetivo del catalogador, así como también las molestias de los lenguajes artificiales.

Aún no se ha puesto en práctica ningún sistema. Sin embargo, numerosos estudios se hallan en curso en ese sentido. Aquí examinaremos las dos principales maneras de abordar la cuestión: la manera estadística y la lingüística. Atacando el problema con ayuda de la estadística, BORKO ha desarrollado una técnica empírica mediante la cual se puede construir un plan de clasificación sumaria, construido de manera semi-automática al tratar el lenguaje original de los documentos. Se ha calculado la frecuencia de todas las palabras en una serie de documentos y se han seleccionado las 90 palabras-tipo más empleadas para un análisis ulterior. Se pudo establecer una matriz de 90 términos con 618 documentos. Entonces, el grado de asociación entre los términos fué calculado en función del número de veces que aparecían en el mismo grupo de documentos. Se logró una matriz "término-término". Un análisis de los elementos reveló entonces cuáles palabras se hallan emparentadas entre si para formar un grupo y éstos son interpretados como categorías de clasificación para el agrupamiento de los documentos tratados. Aún no ha sido comprobada la estabilidad de esas categorías ni la posibilidad de su aplicación en cuanto a la eficacia del trabajo de selección. (11).

Entre las otras técnicas matemáticas de elaboración de esquemas de clasificación, podemos citar la "clump theory" desarrollada por Parker Rhodes del Grupo de Investigaciones Lingüísticas de Cambridge, Inglaterra (12), el análisis de las clases laten

tes propuesto por Barker (13) y las ecuaciones Bayesianas de previsión, utilizadas por Maron (14).

La computadora ha sido empleada en esas experiencias para tratar vastos volúmenes de texto en lengua natural. Esos investigadores sostienen que ciertos sistemas de clasificación empírica podrían resultar más eficaces en la clasificación y selección de informaciones que los métodos tradicionales.

Se ha propuesto el análisis estadístico de los textos no sólo para crear categorías y atribuirles a los documentos, sino también para organizar esas categorías. Doyle ha preconizado la construcción de "mapas ruterros" semánticos que servirían como Léxicos gigantes elaborados automáticamente y que comprendería todas las palabras realmente halladas en una biblioteca y empleadas por los autores.

El análisis del lenguaje natural de una biblioteca de documentos podría producir a través de ese medio una organización de las categorías correspondientes a las asociaciones indicadas por algunos autores. La representación bajo forma de diagramas o de "mapas ruterros" de las asociaciones textuales se basaría, pues, en el análisis de las co-apariciones de palabras. Mas la utilidad de tales mapas ruterros se halla todavía lejos de haber sido precisada (15).

Se han llevado a cabo numerosos intentos, con variado éxito, para catalogar los documentos mediante computadoras. Montgomery y Swanson afirmaron que una computadora puede asignar los rubros de clasificación con gran exactitud. Ellos estudiaron varios miles de entradas en una bibliografía, el Index-Medicus, encontrando que una proporción considerable de las entradas comprendía palabras sinónimas o idénticas a las de los rubros correspondientes, atribuidas sin el concurso de máquinas. Sacaron como conclusión que el 80% del trabajo de clasificación habría podido ser realizado sin errores por una computadora, trabajando sólo sobre los títulos, siempre que se proporcionase a la máquina una lista de sinónimos apropiados y de asociaciones semánticas para cada rubro. (16).

Luhn y Baxendale han tratado de establecer estadísticamente los descriptores de un documento, basándose en la frecuencia de empleo de las palabras en el documento. Resulta, por consiguiente, posible establecer un número predeterminado de las palabras más usadas. Estas palabras representarían el calco mecanizado de los esquemas de clasificación a base de palabras, tales como los "uniterms". Luhn ha propuesto el empleo complementario de un Léxico, a efectos de eliminar los sinónimos y de prever las relaciones genéricas (17).

Baxendale ha obtenido estadísticas de distribución y de frecuencia de las palabras estudiando ciertas partes de textos más bien que textos íntegros.

Su análisis se basa en la supresión de los elementos verbales del género conjuntivo o reiterativo, en la selección de la primera y la última frase solamente, en cada párrafo del documento y en el tratamiento de las cuatro primeras palabras que siguen a cada preposición. (18).

Se han redactado resúmenes analíticos con medios automáticos, empleando técnicas similares. La técnica de auto-abstracting de Luhn requiere la alimentación de la computadora con el texto íntegro, y el cálculo de distribución y frecuencia de las palabras, omitiéndose las de significado poco importante. Las palabras de gran frecuencia son devueltas a su frase de origen y anotadas. Para cada frase se calcula "una tasa de significado" que refleja la frecuencia de palabras significativas en cada frase y su proximidad. Cuando se ha obtenido un valor numérico en todas las frases, se las clasifica y se elige a las que posean el más elevado para formar el resumen analítico. Luhn afirma que "allí donde el máximo de palabras diferentes y de empleo frecuente se encuentren más próximas, mayor es la probabilidad de que la información transmitida sea más representativa del artículo" (19).

Oswlad (20), Edmundson y Willys han propuesto otros criterios de significación para los grupos de palabras y las frases (21).

La aproximación estadística se halla todavía en la etapa experimental. La investigación actual trata de encontrar sobretodo categorías de clasificación, trata de organizar esas categorías y de aplicarlas a los documentos. Existe alguna confusión entre estos diferentes objetivos. Por lo demás, todavía no se ha examinado la eficacia relativa del tratamiento de las asociaciones estadísticas entre palabras significativas en los documentos en relación con el análisis del contenido humano y subjetivo. A primera vista, parecería que esta solución puede al menos igualar el trabajo de los catalogadores humanos. No disponemos de ninguna evidencia en cuanto a la eficacia de una catalogación basada sobre el análisis estadístico en la selección misma de la información.

En efecto, muchos investigadores tratan de determinar hasta qué punto son aplicables las técnicas del análisis lingüístico a la selección de la información. Harris ha discutido sobre las posibles ventajas de transformación lingüísticas y sostiene que el método torna posible la reducción de un texto a una serie de frases-clase que contengan más o menos toda la información del texto original. (22). Salton ha hecho un examen del análisis sintáctico mecanizado representando las correlaciones sintácticas por medio de árboles (23).

Melton se halla estudiando -de manera cuasi lingüística- la posibilidad del tratamiento automático de resúmenes analíticos ya redactados. El buen método exige que se trate automáticamente a los textos impresos de resúmenes analíticos sobre metalurgia, a un nivel más elevado que la simple consulta de un diccionario; más exactamente, es preciso localizar los términos metalúrgicos en el texto y encontrar ciertos símbolos para representar las relaciones entre los términos considerados necesarios para la selección. El análisis no es puramente gramatical. Sólo se recurrirá a la gramática en los casos en que la estructura de la frase lo vuelva necesario.

IV. LA ORGANIZACION DE FICHEROS. -

La mayoría de los ficheros se hallan organizados sobre la base de relaciones lógicas. Corrientemente se emplea una gran número de sistemas de clasificación. "a priori". La asignación de términos a los documentos da lugar a la aparición de un gran número de sub-classes, y las relaciones entre término y documento brindan un instrumento

útil para la organización de los ficheros.

Se han utilizado matrices de asociaciones para calcular el grado de relación entre documentos de modo de poder identificar por ese medio agrupaciones de documentos. Maron y Kuhns han propuesto algunas técnicas para agrupar automáticamente los documentos basándose en las relaciones estadísticas de los términos del catálogo.

Una matriz "términos-documentos" se construye con pesos atribuidos a cada entrada del catálogo. Partiendo de la matriz, podemos obtener medidas de distancia entre documentos en relación con su grado de semejanza, permitiendo cada una de esas medidas un agrupamiento de los documentos según sus semejanzas (25).

Stiles ha propuesto una técnica para grupamiento similar de términos de catálogo con una matriz de relaciones "término-documento", como base. Se redacta una lista de términos partiendo del número de veces que los mismo aparecen en asociación con un término dado. De este modo se obtiene a partir de una colección de documentos ciertos "perfiles de términos". El cálculo de los elementos de asociación prevee el agrupamiento de documento sobre la base de sus relaciones mutuas (26).

La función de una organización lógica de los ficheros consiste en agrupar los documentos que se asemejan a fin de permitir la selección de los documentos ligados. Los documentos en el fichero son agrupados según la aptitud que se les ha previsto para satisfacer a los pedidos de búsquedas. Gofman y Badger han desarrollado un sistema de organización de ficheros mediante computadoras, fundándose en la historia de la utilización anterior de ficheros. De igual modo, se puede modificar un fichero según las cuestiones planteadas (27).

El desarrollo de los "índices de citas" muestra un esfuerzo por disponer los ficheros según la genealogía de los documentos. A partir de un documento citado podremos hallar una lista de sus vástagos. Y aún es posible buscar a los hijos de los hijos, y así sucesivamente hasta los últimos artículos que hayan sido incluidos. En sí mismo, un índice de citas constituye un medio de encontrar la huella de los descendientes (28); por lo demás es empírico más bien que limitado por el lenguaje. De igual modo se lo puede considerar como un sistema que se organiza solo, en la medida en que cada vez que se introduce un documento en el catálogo, la literatura anterior es reclasificada según su empleo actual. La organización viene, pues, a posteriori más bien que a priori. Esas investigaciones fueron emprendidas por Garfield en el Institute of Scientific Communication, en materia de genética (29); y por Tukey, del Statistical Techniques Research Group, de Princeton (30). El lugar que ocupará la automatización en los catálogos de citas no ha sido aún determinado.

V. SISTEMAS AUTOMATICOS DE DIFUSION. -

En 1959, Luhn propuso un sistema mecanizado para difundir la información entre los usuarios mediante el empleo de computadoras (31). Desde esa fecha el sistema de la S.D.I. (Selective Dissemination of Information) ha sido desarrollado y ampliado en muchos casos, pero no de manera muy extensa fuera del IBM (32)

Para cada usuario se establece un perfil de interés, que esboza y define los centros de interés del mismo.

El perfil se compone de términos que figurarán verosimilmente en los documentos apropiados. Esos perfiles se conservan en cinta magnética para su procesamiento por la computadora. A medida que va recibiendo los documentos, un personal sin calificación técnica especial prepara y luego perfora los resúmenes analíticos. Las palabras-clave son extraídas de ellos por la computadora. Las palabras-clave que representan documentos son comparadas con los perfiles de interés guardados en cinta magnética. Cada vez que los términos del usuario y de un documento coinciden, se envía el resumen analítico al usuario. Este, puede elegir su perfil y aún balancear los términos buscados para determinar el "hit level", que mida el grado de similitud deseada entre el perfil y los documentos comparados. Los perfiles pueden ser modificados automáticamente, según las reacciones del usuario, sobre la pertinencia de los documentos brindados. Resulta evidente que la pertinencia de los documentos brindados y difundidos no guarda ninguna relación con la técnica de su difusión. La comparación de términos documentales y de perfiles de interés se halla sujeta a los mismos problemas de pertinencia encontrados en los sistemas de investigación descriptos más arriba.

VI. CONCLUSIONES. -

Nos parece innegable que la mecanización de numerosos sistemas de segundo orden ya ha sido realizada sea en la práctica, sea en su etapa de investigación. En efecto, la selección de las informaciones por medio de computadoras ya ha sido puesta en práctica en muchos casos. Se ha logrado más de lo que parece por el hecho de que la estructura fundamental de los sistemas de selección ya ha sido esbozada. La tentativa de mecanizar la búsqueda de las informaciones ha puesto en evidencia que ésta no es sino un sistema secundario en relación con el sistema total.

Por eso no puede sorprendernos el hecho de que el enfoque tecnológico del problema haya conocido tan extraordinario favor, y que los esfuerzos se hallen concentrados actualmente en la construcción, mejoramiento y automatización de los sistemas secundarios. El éxito por ese camino tal vez conduzca algún día a la mecanización de los sistemas de selección en su totalidad. Es preciso insistir en el hecho de que el empleo de una computadora nada tiene que ver con la eficacia de un sistema de selección. Sin embargo, la mecanización contribuirá indirectamente a la solución del problema, pues esa solución es función a la vez de la competencia y de la eficacia.

Alan M. Rees.

Assistant Director
Centre for Documentation and
Communication Research School
of Library Science Western
Reserve University Cleveland
6, Ohio, U. S. A.

REFERENCIAS:

1. Bush, Vannevar, "As we may think", Atlantic Monthly, July 1945, v. 176, p. 101.
2. Dennis, B.K., J.J. Brady and J.A. Dovel Jr., "Five Operational Years of Inverted Index Manipulation and Abstract Retrieval by an electronic computer", Journal of Chemical Documentation, v. 2, oct. 1962, p. 234-242.
3. Automation of ASTIA-1959, A preliminary Report, dec. 1959, 50 p., AD 227.000; Automation of ASTIA-1960, Dec. 1960, 32 p. AD. 247.000; Evolution of the ASTIA Automated Search and Retrieval System, A Report Bibliography, Jan. 1961, 66 p., AD. 252.000.
4. Perry, J.W. and A. Kent, Tools for Machine Literatura Searching, Interscience, New York, N. Y. 1958; Information Retrieval in Action-Western Reserve University Press, 1963; Hyslop, M. R. Machine Literature Searching-From Experiment to Experience, American Documentation, v. 12, Jan. 1961, p. 49-52.
5. Progress Report on the MEDLARS Project in Proceedings of the Second International Congress of Medical Librarianship, Bulletin of the Medical Library Association, v. 52, Jan. 1964.
6. Milligan, J. L. "Information Retrieval in a Technical Library" talk given at the 15th National Conference of the Association for Computing Machinery, Milwaukee, Aug. 1960, p. 148-177.
7. Costello, John C., "Storage and Retrieval in a Tecnical Library" and Patent Information by Links and Roles in Dupont", American Documentation, n^o 2, 1961, p. 111-120.
8. Brenner, E. H., Weil, B. H. and Rawson, N. E., "A New Cwntralized Information Retrieval System for the Petroleum Industry Including a computer search system and two Manual Indexes", in Automation and Scientific Communication, v. 2 American Documentation Institute, Oct. 1963, p. 291-292.
9. Verhoeff, J., W. Goffman and Jack Belzer, "Inefficiency of the Use of Boolean Functions for Information Retrieval Systems", Communications of the Association for Computing Machinery, v. 4, Dec. 1961, p. 557-559.
10. Maron, M. E. and J. L. Kuhns, "On Relevance, Probabilistic Indexing and Information Retrieval", Jornal of the Asociati6n for Computing Machinery, v. 7, Jul, 1960, p. 216-244.
11. Borko, H., "Research in Document Classification and File Organization", System Development Corporation, SP. -L423, Nov. 13-1963, 12 p.
12. Parker-Rhodes, A. F., "Contribution to the Theory of Clumps", Cambridge Language Research Unit, Cambridge, Cambs, 1961.

3. Baker, B. B. "Information Retrieval Based upon Latent Class Analysis", Journal of the Association of Computing Machinery, v.9, Oct.1962, p.512-521.
4. Maron, M. E., "Automatic Indexing: An Experimental Inquiry", Journal of the Association for Computing Machinery, v.8, Jul.1961, p.407-417.
5. Doyle, L. B., "Semantic Road Maps for Literature Searchers", Journal of the Association for Computing Machinery, v.8, Jul.1961, p.553-578.
6. Montgomery, C. and D. R. Swanson, "Machine-Like Indexing by People", American Documentation, v.13, Oct.1962, p. 359-366.
7. Luhn, H. D., "A Statistical Approach to Mechanized Encoding and Searching of Literary Information", IBM Journal of Research and Development, v.1, Oct.1957, p.309-317.
8. Baxendale, P. "Machine-made Index for Technical Literature an Experiment", IBM Journal of Research and Development, v.2, Oct.1958, p.354-361.
9. Luhn, H. D., "Automatic Creation of Literature Abstracts", IBM Journal of Research and Development, v.2, Apr.1958, p.159-165.
10. Oswald, V. A., "Automatic Indexing and Abstracting of the Contents of Documents, Oct. 31, 1959, Rome Air Development Center, RADC-TR-59-208.
11. Edmundson, H. P. and R. E. Willys, "Automatic Abstracting and Indexing-Survey and Recommendations", Communication for the Association of Computing Machinery, v.4, May. 1961, p.225-234.
12. Harris, H. S., "Linguistic Transformations for Information Retrieval", International Conference on Scientific Information Proceedings, National Academy of Sciences-National Research Council,
13. Salton, G. "Manipulation of Trees in Information Retrieval", Communications of the Association for Computing Machinery, p.103-114; Salton, G. and Edward Sussengutt, Jr., "Automatic Structure Matching Procedures and Some Typical Retrieval Applications", Scientific Report No. ISR-4: Information Storage and Retrieval, Section V, Cambridge, The Computation Laboratory, Harvard University, August.1963.
14. Melton, J. M., "Automatic Processing of Metallurgical Abstracts for the Purpose of Information Retrieval", Interim Report, NSF-2, Feb.1964.
15. Maron and Kuhns, Op.cit., 10 above.
16. Stiles, H. E., "The Association Factor in Information Retrieval", Journal of the Association of Computing Machinery, v.8, Apr. 1961, p.271-279.
17. Badger, G. F. Jr., and William Goffman, "An Experiment with File Ordering for Information Retrieval", (to be published).

Tukey, John, W., "The Citation Index and The Information Problem. Opportunities and Research in Progress", Statistical Techniques Research Group, Princeton University, 1963, p. 6

Garfield, E., "Citation Indexes for Science", *Science*, v.122, 1955, p.108-111; Garfield, E. and I.H. Sher, "New Factors in the Evaluation of Scientific Literature Through Citation Indexing", American Documentation, v.17, July 1963, p.195-201.

Tukey, Op.cit., citation 25 above.

Luhn, H.D., "Selection Dissemination of New Scientific Information with the aid of Electronic Processing Equipment", IBM, Yorktown Heights, New York, 1954.

Kraft, Donald H., "A Selection Dissemination of Information (SDI) System Using an IBM Computer to make users' Interests with Current Abstracts, IBM, Chicago, Nov. 1963.