

From Computational Linguistics to Algorithmic Historiography

Presented at University of Pittsburgh

by

Eugene Garfield

Chairman Emeritus, ISI[®] & Publisher, The Scientist[®]

3501 Market Street, Philadelphia, PA 19104

Tel. 215-243-2205, Fax 215-387-1266

email: garfield@codex.cis.upenn.edu

Home Page: www.EugeneGarfield.org

Lazerow Lecture held in conjunction with panel on "Knowledge and Language: Building Large-Scale Knowledge Bases for Intelligent Applications"

September 19, 2001

Abstract

This Lazerow Lecture was presented to honor the memory of Professor Casimir Borkowski. It coincided with the inauguration of the Casimir Borkowski Scholarship. The introduction is autobiographical and traces our accidental encounter in 1949, after our graduation from Columbia University that year. Our forty-year friendship continued through graduate school with digressions until receiving doctorates at the University of Pennsylvania under Zellig Harris, Casimir in 1958 and myself in 1961.

In the fifties, Cas worked for Leon Dostert at Georgetown University in the project involving mechanical translation of Russian to English. After leaving Penn he worked at IBM in Yorktown Heights and then taught at the University Pittsburgh both in the computer science and information science programs.

After commenting on the MT problem, the talk segues to work on algorithmic historiography and its relation to Information Retrieval. My work on historical networks began with Irv Sher in 1964 in a project which produced a map showing the development of DNA from Mendel to Niremberg in 1962. The DNA map is the model for future automatic generation of historical or genealogical maps of papers or topics. The technique is illustrated by the example of bibliographic coupling as first enunciated by MM Kessler in 1963. The map is generated by doing a search of the *Science Citation Index[®]* on the *ISI Web of Science[®]* or CD ROM and generating an Export file of source records, which is then processed with newly developed software called "histcomp". The software creates a chronological file of nodal papers and a list of non-nodal papers, patents and books. From these the milestone papers of the field are easily identified based primarily on citation frequency either within the nodal bibliography or in the outside global universe of the WOS domain. An editing module identifies variations and errors in cited references which can then be edited and integrated into the main matrix. The edited matrix is in turn used to generate maps which help to visualize the genealogical flow of the field from primordial papers to the present.

In the summer of 1949, I spent a month hitchhiking in Canada with my younger brother Ralph. By chance, we met Casimir Borkowski on a local bus in Quebec. Like me, he had just graduated from Columbia University. We traveled together on a long journey through the Laurentians up to Shipshaw Dam, down the Saguenay River trail, around the Gaspé Peninsula to New Brunswick, Prince Edward Island, and Nova Scotia. All on a budget of \$100.

After we returned to New York City, Cas introduced me to his mother Helen and his stepfather, a Polish diplomat, who soon returned to Poland. Mrs. Borkowski had been an actress in Poland. In 1939, she and Cas fled the Nazis first to Paris, then to Marseilles, and then to Algeria where they spent the rest of the war. They came to America via Lisbon. I believe Cas majored in dramatic arts but his real love was language and was fluent in Polish, Russian, French, German, and Spanish. About 1950, he went to Georgetown University to work with Leon Dostert, a leader in mechanical translation.

After a few laboratory jobs, I wound up working for San Larkey at the Johns Hopkins Welch Medical Library Indexing Project. After two years I returned to Columbia to attend the School of Library Service. After graduation, in 1954, I planned to join Cas at Georgetown but due to financial difficulties, I joined Smith Kline and French Laboratories in Philadelphia temporarily as a documentation consultant. By the time 1954 ended, however, the Georgetown Project was on the wane and shortly afterward, Cas came to Philadelphia to work on his Ph.D. in structural linguistics with Zelig Harris at the University of Pennsylvania.

In 1955 or 1956, Cas introduced me to Professor Harris. The three of us met for lunch. I explained the field of information retrieval (IR) to him since he had never heard about it. The possibility that the National Science Foundation (NSF) might fund a linguistics-based IR project had never occurred to him. I knew that Helen Brownson of NSF would be interested because of her interest in IR and language. When the Welch Project folded, she had encouraged me to find a new academic affiliation since NSF could not support private individuals.

Based on my discussions with Cas and Harris, I applied for a small NSF grant of \$10,000 in April, 1956 to study the application of structural linguistics to information retrieval and listed Professor Harris as the principle investigator and Cas as consultant. Casimir was among a small group of graduate students at Penn that had included Noam Chomsky and Naomi Sager who were involved in transformational grammars. Naomi is now at the New York University Courant Institute of Mathematical Sciences. Chomsky went to MIT where he remains today one of the most-cited scholars of the twentieth century.

My proposal letter to NSF led to a meeting with Harris and Helen Brownson. However, instead of awarding me a \$10,000 grant, she gave Harris a grant of \$500,000 to establish the Transformation and Discourse Analysis Project.¹

By 1958, Harris was sufficiently involved with information retrieval that he presented a paper at the International Conference on Scientific Information in Washington. His topic was “Linguistic Transformation for Informational Retrieval.”² On page 931, in Footnote 1, he acknowledges the support of NSF which also co-sponsored the conference. Helen Brownson referred to his paper in a 1960 report in *Science* about “Research on Handling Scientific Information.”³

In New York in 1954 I had started to put together a Ph.D. program at Columbia. A few professors understood my interests in computerized scientific documentation and agreed to serve on an interdisciplinary committee.

But I could never get them to meet because they were so far apart on the campus. And financial support was unobtainable since I was unaffiliated.

When I told all this to Harris, he agreed to transfer 30 of my graduate credits towards a doctorate in linguistics at Penn. This was only half my earned credits. Since I was now self-employed, I agreed to pay tuition for 30 additional credits. My coursework included participation in seminars on transformational grammar.

In 1959, I became involved in a chemical information project with the U.S. Patent Office sponsored by the Pharmaceutical Manufacturers Association. This eventually led to my thesis topic, “An Algorithm for Translating Chemical Names to Molecular Formulas.” My half-page communication in *Nature* was entitled “Chemico-Linguistics: Computer Translation of Chemical Nomenclature.”⁴ I had told Harris that not everyone agreed that the goal could be accomplished. Since he was not competent to judge its chemical significance, he agreed to have Professor Alan Day, Chair of the Chemistry Department at Penn be the resident expert.

By 1958, Cas Borkowski had finished his Ph.D. His doctoral dissertation was “Transformations of Polish.”⁵ After graduation, he was hired by IBM at Yorktown Heights, NY. During one of my visits with Cas at IBM, he introduced me to Gilbert W. King, then Director of Research. Manfred Kochen was working at IBM about this time but we did not meet until several years later. For those who are not familiar with Kochen’s work, see my talk at the celebration held on the 10th anniversary of his death.⁶ Like Cas, he was a European émigré and died much too early in a productive life.

Cas also introduced me to Phyllis Baxendale who, among other IBM collaborators, published with Wilf Lancaster on vocabulary control.⁷ She had also worked with Peter Luhn in the late fifties.⁸ Luhn was one of the pioneers of auto-abstracting and keywords in context indexes. We first met at the Welch Project in 1952.

Borkowski’s published work reflects some of his academic interests. But much of his work was reported in unpublished technical reports. But these contributions do not give an adequate impression of his impact on the computational linguistics and information science communities. His work in the sixties on algorithmic systems for identifying personal names was truly a pioneering effort which anticipated our preoccupation today with full-text searching on the web.

Every science community consists of a variety of individuals. Some contribute to the advancement of knowledge through experimentation and invention. Others provide thoughtful commentary on the work of their colleagues locally and to the wider international community of

scholars. As I indicated, Cas was involved in a variety of research investigations in information retrieval but his impact on the field was mainly through his role as reviewer, advisor, and teacher.

His speaking knowledge of languages and deep understanding of structural linguistics and computer science, made him an excellent sounding board. During our long friendship, we spent endless hours discussing topics of mutual interest. Cas taught me more about linguistics than I ever learned in school.

After Helen Borkowski died, Cas and I were out of touch for some time. It was only when I was preparing these remarks that I realized that he had also participated in some interesting library-based user surveys. During his tenure at the University of Pittsburgh he worked with Professor MacLeod, now at the University of Florida. He also taught courses both in computer science and library science.

I've said very little about his early work with Leon Dortert on the mechanical translation of Russian. Since those early days, fifty years ago, enormous strides have been made in mechanical translation. But anyone who uses these MT systems knows we have a long way to go. The task has been made easier by the availability of low-cost memory but clearly existing methods like Systran only go so far. In all probability Bar-Hillel's prediction a half century ago that complete automatic translation is impossible, appears to hold up.⁹ Thanks to the remarkable and informative website of John Hutchins in the UK (<http://ourworld.compuserve.com/homepages/WJHutchins/>), we have an excellent description of the milestones of mechanical translation. John refreshed my memory on the work of Yehoshua Bar-Hillel whom I met at MIT during a visit there in 1953. I was introduced to him by James Perry on the same day that he introduced me to Norbert Wiener, the father of cybernetics.

Bar-Hillel's evaluation of mechanical translation is tersely expressed by the following statement:

"Mechanical translation is an instance of 'a well-known situation where accuracy may be traded for speed, and vice versa.' For Bar-Hillel it was already 'obvious' that 'fully automatic MT, i.e. one without human intervention..[was] achievable only at the price of inaccuracy.' The major obstacle to fully automatic translation was that there were no obvious methods "by which the machine would eliminate *semantical ambiguities*.

"However, he stressed (in words which are as valid today as when he was writing) that 'with a lowering of the target, there appear less ambitious aims, the achievement of which is still theoretically and practically viable."

Thanks to John Hutchins, I was able to determine that the 1951 paper by Bar-Hillel in *American Documentation* was actually published in 1953, an important fact when you are trying to determine the exact history of this working paper which was used as the basis for the 1952 conference held at MIT. I've delved into MT simply by way of showing you that in its early history computational linguistics and information retrieval were not so far apart.

When you consider that it took 20 years for the journal *Computational Linguistics* to appear in 1974, we are indeed talking about ancient history when we hark back to the early 50s.

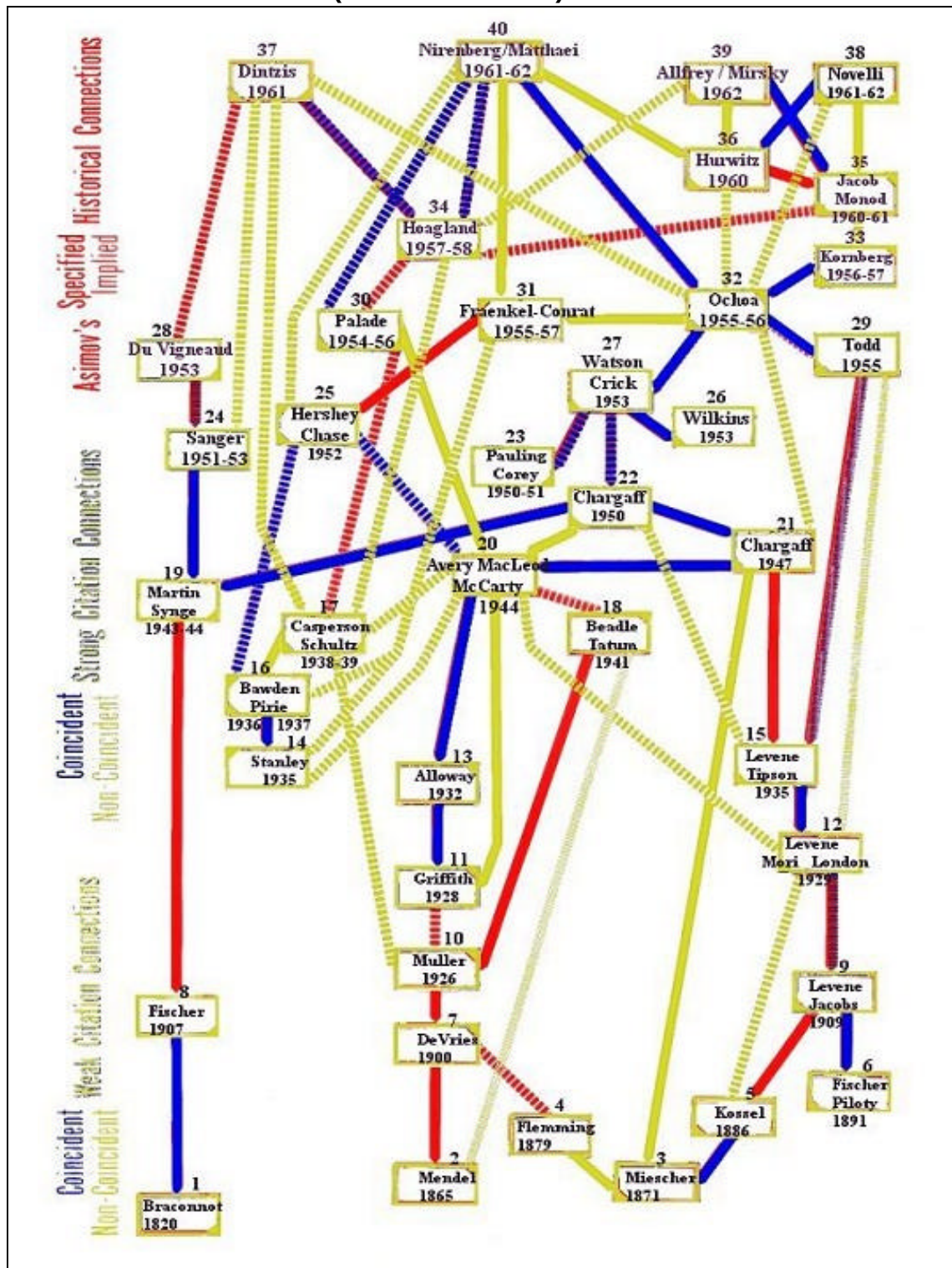
As an aside, I would remind you that full-text natural language searching did not begin with the internet. John O'Connor was doing manual full-text searching research when I met him in the 50s. And Gerry Salton's work in the 60s is well known to most information scientists.

However, it is noteworthy that I was unable to find a published paper that uses the term "full-text" in the title until 1971 in connection with searches of law cases. The term came into heavier use when Dialog and Lexis went up in the early 70s when natural language searching became more popular. The term, natural language, was used by Don Swanson in a 1960 *Science* article "Searching Natural Language Text by Computer."¹⁰ From this observation, I've tried to show how easy it was in those days to move from IR to MT and back again to FT. All were hindered by lack of cheap memory.

These linguistic and retrieval problems are closely related to the problems of automatic abstracting and literature review. While a level of automatic indexing and abstracting has been achieved, automatic reviewing is not yet in sight. I reviewed these problems back in 1965^{11,12} when I discussed the automation of citation indexing. Automating citation involves accounting for aposteriori semantic elements that are beyond ordinary full-text procedures. In creating autonomous citation indexes, that is, citation in context as is done by Steve Lawrence,¹³ and proposed long ago by Henry Small,^{14,15,16} we get closer to semi-automatic reviewing. But the type of linguistic analysis required for these functions is still far beyond us. Even spell checkers can do only the most simplistic grammar checking. I addressed this difficulty in 1965 when I spoke about the automatic creation of cited references.^{6,7} Can an automaton select the appropriate references for statements made by the author? Even humans cannot agree on which document an author intended to cite. Nor can they agree on whether a reference is needed.¹⁷ Displaying citing sentences and paragraphs is one thing but integrating these statements into a meaningful integrated review is quite another matter. Even matching an author's name against a list of candidate papers does not guarantee a correct match unless one has certain apriori information.

This leads me to my closing topic -- algorithmic historiography. This was the subject of a report prepared by Irv Sher and myself back in 1964.¹⁸ That report was not then widely disseminated and only recently we posted it on the web.⁹ We showed how citation analysis might be used to facilitate scientific historiography, that is, writing the history of science. In that process we would produce what we call an historiograph. A basic premise here is that most cited works in any field include the key developments as represented by published papers and book.

SLIDE 1: DNA HISTORIOGRAPH (same as Slide 11)



In our study of the history of DNA from Mendel to Nirenberg, we created the citation network for 40 key events and compared the results to a history created by Issac Asimov in his book on the topic.¹⁹ Asimov had an almost perfect memory, but our analysis showed he had omitted some key links.

After a 36-year lapse, I decided to go back to our earlier plan to automate this process. To do this, we start with a collection of papers, each of which is represented by its source record in

Science Citation Index on the Web of Science. The algorithm uses the lists of cited references in the collection to create a mini-citation index of the field, as was done for the history of DNA. The software produces a basic chronological matrix of the bibliography, and then generates listings of most-cited (core) papers in the input network. However, it also produces a list of the most-cited papers outside the main network. By an editing procedure the user can add any of these non-nodal papers to the main group and thereby to the map.

SLIDE 2: WOS RECORD SHOWING CITED REFERENCES

ISI Web of SCIENCE®

Powered by ISI Web of Knowledge_{SM}

HOME

HELP

DATE & DB LIMITS

GENERAL SEARCH

CITED REF SEARCH

SEARCH RESULTS

LOG OFF

Cited References

AN EXAMINATION OF CITATION INDEXES

MARTYN J

ASLIB PROCEEDINGS

17 (6): 184-196 1965

FIND RELATED RECORDS

[Explanation](#)

Clear the checkbox to the left of an item if you do not want to search for articles that cite the item when looking at Related Records.

Cited Author	Cited Work	Volume	Page	Year
☒ *I SCI INF	SCI CIT IND		17	1961
☒ CLEVERDON CW	REPORT TESTING ANAL			1962
☒ GARFIELD E	SCIENCE	144	649	1964
☒ GREENWOOD JA	ANN MATH STATISTICS	1		1962
☒ HAMMERSLEY JM	NATURE	202	330	1964
☒ KEEN EM	ASLIB P	16	246	1964
☒ KEENAN S	J LIT PHYSICS			1964
☒ KESSLER MM	AM DOC	14	10	1963
☒ LIPETZ B	COMPILATION EXPT CIT			1961
☒ LIPETZ B	EVALUATION IMPACT CI			1964
☒ MARTYN J	J DOC	20	212	1964
☒ MARTYN J	REPORT INVESTIGATION			1964
☒ PRICE DJD	DEC S HIST REL SCI T			1964
☒ RESNICK A	SCIENCE	134	1004	1961
☒ TERRY JE	J DOCUMENTATION	21	139	1965
☒ TUKEY JW	1962 PRINC U STAT TE			
☒ URQUHART DJ	J DOC	15	21	1959
☒ WALDHART TJ	THESIS U WISCONSIN			1964

A search of the *Web of Science* produces this type of source record. The search process is completed and a marked list is created.

SLIDE 3 TYPICAL EXPORT RECORDS FROM THE WEB OF SCIENCE

```

FN ISI Export Format
VR 1.0
PT Journal
AU MARTYN, J
TI AN EXAMINATION OF CITATION INDEXES
SO ASLIB PROCEEDINGS
NR 18
CR *I SCI INF, 1961, SCI CIT IND, P17
  CLEVERDON CW, 1962, REPORT TESTING ANAL
  GARFIELD E, 1964, SCIENCE, V144, P649
  GREENWOOD JA, 1962, ANN MATH STATISTICS, V1
  HAMMERSLEY JM, 1964, NATURE, V202, P330
  KEEN EM, 1964, ASLIB P, V16, P246
  KEENAN S, 1964, J LIT PHYSICS
  KESSLER MM, 1963, AM DOC, V14, P10
  LIPETZ B, 1961, COMPILATION EXPT CIT
  LIPETZ B, 1964, EVALUATION IMPACT CI
  MARTYN J, 1964, J DOC, V20, P212
  MARTYN J, 1964, REPORT INVESTIGATION
  PRICE DJD, 1964, DEC S HIST REL SCI T
  RESNICK A, 1961, SCIENCE, V134, P1004
  TERRY JE, 1965, J DOCUMENTATION, V21, P139
  TUKEY JW, 1962 PRINC U STAT TE
  URQUHART DJ, 1959, J DOC, V15, P21
  WALDHART TJ, 1964, THESIS U WISCONSIN
BP 184
EP 196
PG 13
JI Aslib Proc.
PY 1965
VL 17
IS 6
GA CDN98
J9 ASLIB PROC
UT ISI:A1965CDN9800001
ER

```

The output of the marked list is connected to the export format which contains tags for all the informational elements.

As shown in this slide, we have a typical export record from the WOS.

SLIDE 4: CHRONOLOGICAL FILE

[Outer nodes](#) [Missing links?](#) [Journal list](#) [All-Author list](#)

Citations to Kessler's Bibliographic Coupling and papers with BC in title/abstract
Nodes: 223

Sorted by **year, journal, volume, page**.

Cited nodes	Nodes / Authors	GCS	LCS
0	0 1963 AMERICAN DOCUMENTATION 14(1):10-& KESSLER MM <i>Bibliographic Coupling Between Scientific Papers</i>	128	134
1	1 1963 AMERICAN DOCUMENTATION 14(4):289-& GARFIELD E <i>Citation Indexes in Sociological and Historical Research</i>	61	5
0	2 1963 IEEE TRANSACTIONS ON INFORMATION THEORY 9(1):49-& KESSLER MM <i>An Experimental Study of Bibliographic Coupling Between Technical Papers</i>	8	8
1	3 1963 INFORMATION STORAGE AND RETRIEVAL 1(4):169-187 KESSLER MM <i>Bibliographic Coupling Extended in Time - 10 Case Histories</i>	14	15
1	4 1964 ASLIB PROCEEDINGS 16(2):48-63 [Anon] <i>ASLIB 37th Annual Conference - University of St Andrews, 24th- 26th September 1963</i>		0
1	5 1964 ASLIB PROCEEDINGS 16(4):132-152 LANCASTER FW <i>Mechanized Document Control - A Review of Some Recent Research</i>	3	0
2	6 1964 ASLIB PROCEEDINGS 16(8):246-251 KEEN EM <i>Citation Indexes</i>	5	2
0	7 1964 JOURNAL OF DOCUMENTATION 20(4):236-236 MARTYN J <i>Bibliographic Coupling</i>	12	7
1	8 1964 NACHRICHTEN FUR DOKUMENTATION 15(3):122-130 MODEL F <i>Citation Index and Retrospective Cataloging - Examples of Citation Documentation</i>	5	0
1	9 1964 SCIENCE 144(361):649-& GARFIELD E <i>Science Citation Index - New Dimension in Indexing - Unique Approach Underlies Versatile Bibliographic Systems for Communicating and Evaluating Information</i>	92	19

The randomly organized export file is sorted by the software to produce a precise chronological file.

In next slide, we have a piece of the chronological file showing all papers that have cited M. M. Kessler or have used the term Bibliographic Coupling in their titles.

SLIDE 5: MOST-CITED PAPERS FROM THE STARTING BIBLIOGRAPHY

[Outer nodes](#) [Missing links?](#) [Journal list](#) [All-Author list](#)

Citations to Kessler's Bibliographic Coupling and papers with BC in title/abstract

Nodes: 223

Sorted by **LCS**.



Cited nodes	Nodes / Authors	GCS	LCS
0	0 1963 AMERICAN DOCUMENTATION 14(1):10-& KESSLER MM <i>Bibliographic Coupling Between Scientific Papers</i>	128	134
3	75 1973 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE 24(4):265-269 SMALL HG <i>Cocitation in Scientific Literature - New Measure of Relationship Between 2 Documents</i>	235	82
3	10 1965 AMERICAN DOCUMENTATION 16(3):223-233 KESSLER MM <i>Comparison of the Results of Bibliographic Coupling and Analytic Subject Indexing</i>	42	43
2	15 1965 PHYSICS TODAY 18(3):28-& KESSLER MM <i>MIT Technical Information Project</i>	36	39
2	88 1974 SCIENCE STUDIES 4(1):17-40 SMALL HG; GRIFFITH BC <i>Structure Of Scientific Literatures . 1. Identifying And Graphing Specialties</i>	212	37
1	78 1973 NAUCHNO-TEKHNICHESKAYA INFORMATSIYA SERIYA 2- INFORMATSIONNYE PROTSESSY I SISTEMY 2(6):3-8 MARSHAKOVA IV <i>System Of Document Connections Based On References</i>	24	22
1	9 1964 SCIENCE 144(361):649-& GARFIELD E <i>Science Citation Index-New Dimension in Indexing - Unique Approach Underlies Versatile Bibliographic Systems for Communicating and Evaluating Information</i>	92	19
4	58 1971 INFORMATION STORAGE AND RETRIEVAL 6(6):417-& SCHIMINOVICH S <i>Automatic Classification and Retrieval of Documents by Means of Bibliographic Pattern Discovery Algorithm</i>	27	16
1	3 1963 INFORMATION STORAGE AND RETRIEVAL 1(4):169-187 KESSLER MM <i>Bibliographic Coupling Extended In Time - 10 Case Histories</i>	14	15

In slide 5, we see the list of papers sorted by citation frequency within the "local," that is, the starting bibliography. Note that LCS means local citation score.

SLIDE 6: SORT BY GLOBAL CITATION SCORE

[Outer nodes](#) [Missing links?](#) [Journal list](#) [All-Author list](#)

Citations to Kessler's Bibliographic Coupling and papers with BC in title/abstract

Nodes: 223

Sorted by **GCS**.



Cited nodes	Nodes / Authors	GCS	LCS
3 65	1971 MINERVA 9(1):66-100 ZUCKERMAN H; MERTON RK <i>Patterns of Evaluation In Science - Institutionalisation, Structure and Functions of Referee System</i>	275	2
3 75	1973 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE 24(4):265-269 SMALL HG <i>Cocitation in Scientific Literature - New Measure of Relationship Between 2 Documents</i>	235	82
1 27	1967 AMERICAN SOCIOLOGICAL REVIEW 32(3):377-390 COLE S; COLE JR <i>Scientific Output and Recognition - Study in Operation of Reward System in Science</i>	233	9
2 88	1974 SCIENCE STUDIES 4(1):17-40 SMALL HG; GRIFFITH BC <i>Structure of Scientific Literatures .1. Identifying and Graphing Specialties</i>	212	37
0 0	1963 AMERICAN DOCUMENTATION 14(1):10-& KESSLER MM <i>Bibliographic Coupling Between Scientific Papers</i>	128	134
12 111	1981 LIBRARY TRENDS 30(1):83-106 SMITH LC <i>Citation Analysis</i>	97	10
1 46	1969 AMERICAN SOCIOLOGICAL REVIEW 34(3):335-352 CRANE D <i>Social Structure in a Group of Scientists - Test of Invisible College Hypothesis</i>	96	9
1 9	1964 SCIENCE 144(361):649-& GARFIELD E <i>Science Citation Index-New Dimension In Indexing - Unique Approach Underlies Versatile Bibliographic Systems for Communicating and Evaluating Information</i>	92	19
1 149	1987 JOURNAL OF INFORMATION SCIENCE 13(5):261-276 KING J <i>A Review of Bibliometric and Other Science Indicators and Their Role in Research Evaluation</i>	65	4

In Slide 6 we see the sort by Global Citation Score, that is, the total citation frequency for each nodal paper in the SCI.

SLIDE 7: AUTHORS' LIST

Ranked All-Author list

Total: 217

Sorted by **Pubs**

<u>Name</u>	<u>TGCS</u>	<u>TLCS</u>	<u>Pubs</u>
GARFIELD E	197	26	<u>11</u>
SMALL HG	488	125	<u>10</u>
KESSLER MM	228	239	<u>5</u>
SALTON G	85	15	<u>5</u>
KWOK KL	29	8	<u>4</u>
PAO ML	51	5	<u>4</u>
BICHTELER J	25	16	<u>3</u>
BRAUN T	33	1	<u>3</u>
CAWKELL AE	26	6	<u>3</u>
MARSHAKOVA IV	46	24	<u>3</u>
MIYAMOTO S	27	0	<u>3</u>
SAVOY J	14	6	<u>3</u>
SHARABCHIEV YT	10	1	<u>3</u>
VLACHY J	47	3	<u>3</u>
CLEVELAND DB	15	4	<u>2</u>
COLE JR	288	11	<u>2</u>
COLE S	288	11	<u>2</u>
EATON EA	13	10	<u>2</u>
GATRELL AC	19	3	<u>2</u>
JONES WT	7	2	<u>2</u>
KOCHTANEK TR	9	5	<u>2</u>
LANCASTER FW	11	1	<u>2</u>
LICKLIDER JCR	29	0	<u>2</u>
LOGAN EL	6	2	<u>2</u>
MARTYN J	33	11	<u>2</u>
MCCAIN KW	24	1	<u>2</u>
MERTON RK	280	2	<u>2</u>
MIDORIKAWA N	11	1	<u>2</u>
OCONNOR J	13	4	<u>2</u>
OVERHAGE CF	32	0	<u>2</u>
PERITZ BC	15	2	<u>2</u>

The most-published authors related to bibliographic coupling are ranked by number of papers that cite Kessler. Total citations to the papers are shown.

SLIDE 8: JOURNAL LIST

Ranked Journal list

Total: 80

Title	Pubs
JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE	29
SCIENTOMETRICS	21
INFORMATION PROCESSING & MANAGEMENT	13
INFORMATION STORAGE AND RETRIEVAL	13
PROCEEDINGS OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE	8
NAUCHNO-TEKHNICHESKAYA INFORMATSIYA SERIYA 2-INFORMATSIONNYE PROTSESSY I SISTEMY	8
JOURNAL OF DOCUMENTATION	7
LIBRI	5
AMERICAN DOCUMENTATION	5
NAUCHNO-TEKHNICHESKAYA INFORMATSIYA SERIYA 1-ORGANIZATSIYA I METODIKA INFORMATSIONNOI RABOTY	4
ASLIB PROCEEDINGS	4
LIBRARY QUARTERLY	4
ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY	4
PHYSICS TODAY	4
SCIENCE	3
CZECHOSLOVAK JOURNAL OF PHYSICS	3
JOURNAL OF CHEMICAL DOCUMENTATION	3
JOURNAL OF INFORMATION SCIENCE	3
AMERICAN SOCIOLOGICAL REVIEW	3
LIBRARY RESOURCES & TECHNICAL SERVICES	3
PROCEEDINGS OF THE ASIS ANNUAL MEETING	3
LIBRARY TRENDS	3
MINERVA	2
MEDICINA CLINICA	2
COMMUNICATIONS OF THE ACM	2
JOURNAL OF LIBRARIANSHIP	2
NACHRICHTEN FUR DOKUMENTATION	2
AMERICAN PSYCHOLOGIST	2
COMPUTER NETWORKS	2
NAUCHNO-TEKHNICHESKAYA INFORMATSIYA	1

JASIS and *Scientometrics* have published the most papers on this topic.

SLIDE 9: HIGHLY CITED WORKS OUTSIDE ORIGINAL BIBLIOGRAPHY

Web of Science location:

Cited references outside of this network

Total: 6314 (top 30 shown).

LCS	Reference
31	PRICE DJD, 1965, SCIENCE, V149, P510 WoS
26	GARFIELD E, 1955, SCIENCE, V122, P108 WoS
17	GARFIELD E, 1964, USE CITATION DATA WR WoS
17	GARFIELD E, 1979, CITATION INDEXING WoS
16	GRIFFITH BC, 1974, SCI STUD, V4, P339 WoS
14	MARGOLIS J, 1967, SCIENCE, V155, P1213 WoS
14	GARFIELD E, 1972, SCIENCE, V178, P471 WoS
14	SALTON G, 1983, INTRO MODERN INFORMA WoS
14	MORAVCSIK MJ, 1975, SOC STUD SCI, V5, P86 WoS
13	SALTON G, 1963, J ACM, V10, P440 WoS
13	CRANE D, 1972, INVISIBLE COLLEGES WoS
12	SMALL HG, 1977, SOC STUD SCI, V7, P139 WoS
11	NARIN F, 1976, EVALUATIVE BIBLIOMET WoS
10	KWOK KL, 1975, INFORMATION PROCESSI, V11, P201 WoS
10	PRITCHARD A, 1969, J DOC, V25, P348 WoS
10	WHITE HD, 1981, J AM SOC INFORM SCI, V32, P163 WoS
10	KAPLAN N, 1965, AM DOC, V16, P179 WoS
10	LOTKA AJ, 1926, J WASHINGTON ACADEMY, V16, P317 WoS
10	CHUBIN DE, 1975, SOC STUD SCI, V5, P423 WoS
10	GARFIELD E, 1979, CITATION INDEXING IT WoS
10	KUHN TS, 1970, STRUCTURE SCI REVOLU WoS
10	WESTBROOK JH, 1960, SCIENCE, V132, P1229 WoS
9	GARFIELD E, 1980, CURRENT CONTENT 0609, P5 WoS
9	SMALL HG, 1978, SOC STUD SCI, V8, P327 WoS
9	VANRIJSBERGEN CJ, 1979, INFORMATION RETRIEVA WoS
9	NARIN F, 1972, J AM SOC INFORM SCI, V23, P323 WoS
9	PRICE DJD, 1966, AM PSYCHOL, V21, P1011 WoS
8	GARFIELD E, 1970, NATURE, V227, P669 WoS
8	GRIFFITH BC, 1972, SCIENCE, V177, P959 WoS

And in this slide 9 we have a list of highly cited works outside the initial bibliography. This will include not only articles indexed in the *SCI* but also articles, books, and patents not covered in the *SCI* source indexes. For those that are covered in *WOS*, a hotlink is shown which leads to the corresponding entry in the *WOS* cited reference file. The user can decide whether to add these items to the file.

SLIDE 10: MISSING LINKS

Potentially missed citations

9 nodes have citations that may potentially refer to other nodes.

[12](#) 1965 ASLIB PROCEEDINGS 17(6):184-196

MARTYN J

AN EXAMINATION OF CITATION INDEXES

MARTYN J, 1964, J DOC, V20, P212 may refer to [7](#) MARTYN-J-1964-V20-P236

[42](#) 1968 LIBRARY RESOURCES & TECHNICAL SERVICES 12(4):415-&

HUANG TS

EFFICACY OF CITATION INDEXING IN REFERENCE RETRIEVAL

BROWN SC, 1966, PHYSICS TODAY, V19, P60 may refer to [24](#) BROWN-SC-1966-V19-P59
 BROWN SC, 1966, PHYSICS TODAY, V19, P61 may refer to [24](#) BROWN-SC-1966-V19-P59
 BROWN SC, 1966, PHYSICS TODAY, V19, P64 may refer to [24](#) BROWN-SC-1966-V19-P59
 GARFIELD E, 1963, AM DOC, V14, P290 may refer to [1](#) GARFIELD-E-1963-V14-P289
 GARFIELD E, 1964, SCIENCE, V144, P651 may refer to [9](#) GARFIELD-E-1964-V144-P649

[67](#) 1971 PHYSICS TODAY 24(7):28-&

ZUCKERMAN H; MERTON RK

SOCIOLOGY OF REFEREEING

COLE S, 1968, AM SOCIOLOGICAL REV, V33, P412 may refer to [33](#) COLE-S-1968-V33-P397

[102](#) 1978 SCIENTOMETRICS 1(1):9-34

GILBERT GN

MEASURING THE GROWTH OF SCIENCE - REVIEW OF INDICATORS OF SCIENTIFIC GROWTH

COLE S, 1968, AM SOC REV, V33, P297 may refer to [33](#) COLE-S-1968-V33-P397
 GARFIELD E, 1963, AM DOC, V14, P290 may refer to [1](#) GARFIELD-E-1963-V14-P289

[131](#) 1984 NAUCHNO-TEKHNICHESKAYA INFORMACIYA SERIYA 1-ORGANIZACIYA I METODIKA
 INFORMACIIONNOI RABOTY (12):6-11

SHARABCHIEV YT

APPLICATION OF CLUSTER-ANALYSIS IN SCIENTIFIC INVESTIGATIONS

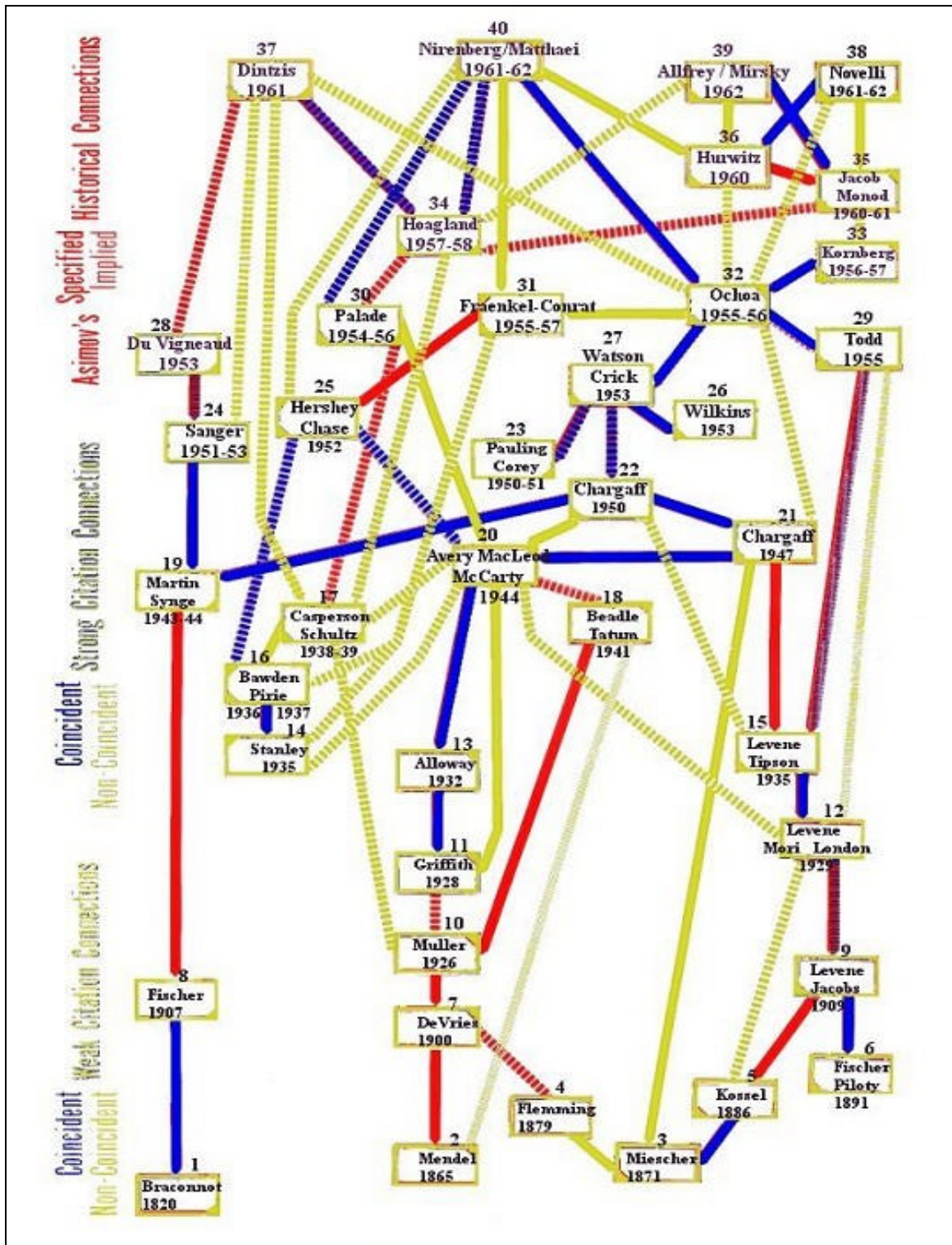
KESSLER MM, 1963, J AM DOC, V14, P99 may refer to [0](#) KESSLER-MM-1963-V14-P10

[138](#) 1985 CZECHOSLOVAK JOURNAL OF PHYSICS 35(12):1389-1436

VLACHY J

An important part of the system involves an error correction routine wherein every doubtful reference is checked against the main file. Errors and variations can be corrected or unified. In the case of the Kessler file we found that citations to Irina Marshakova, the co-discoverer of co-citation clustering, would have been obscured had we not detected the variations in the citations of the Russian journal in which she published.

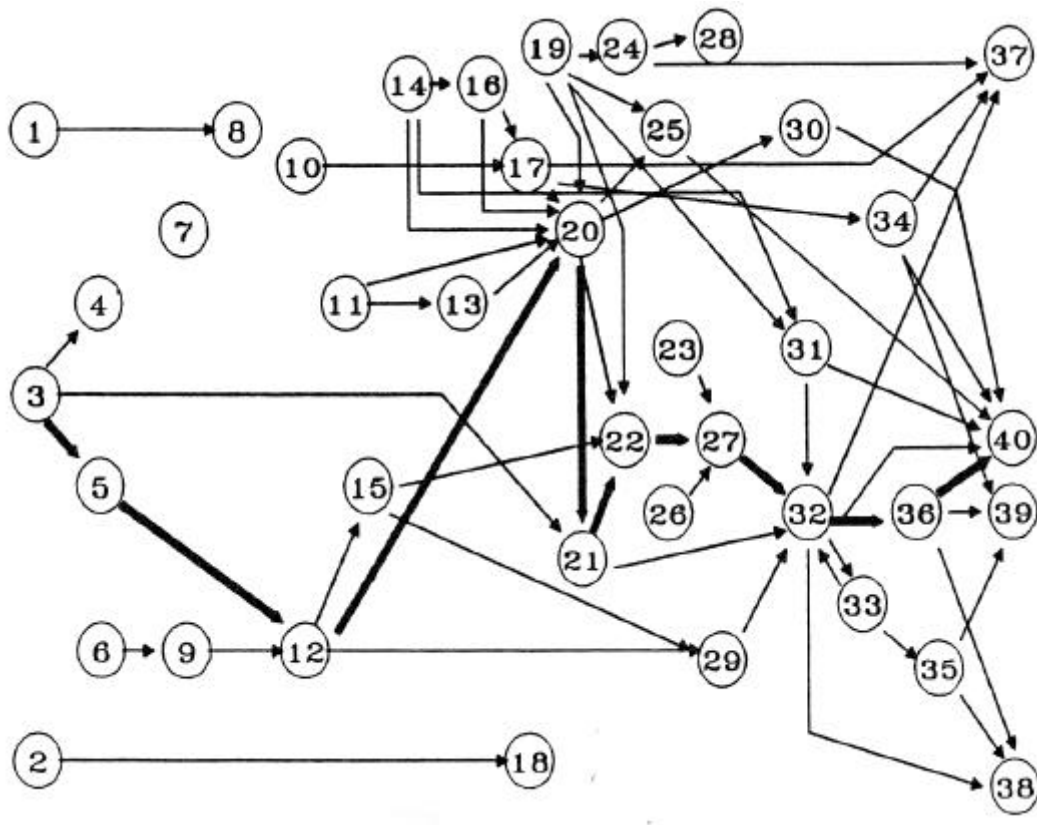
SLIDE 11: HISTORIOGRAPH FROM THE 1964 PROJECT (Same as Slide 1)



In slide 11 we show the original historiograph from the 1964 project. This is the manual prototype of the output for any other topic. However, by using multi-levels of visualization, it is possible to go from this macro view of the field to a more detailed view on a year-by-year basis.

In an extremely fast-moving field, it is essential to preserve if possible, month-by-month outputs of the data. And for precision in recording developments, dates of manuscript submission and acceptance may have to be taken into consideration.

SLIDE 12: HUMMON AND DOREIAN



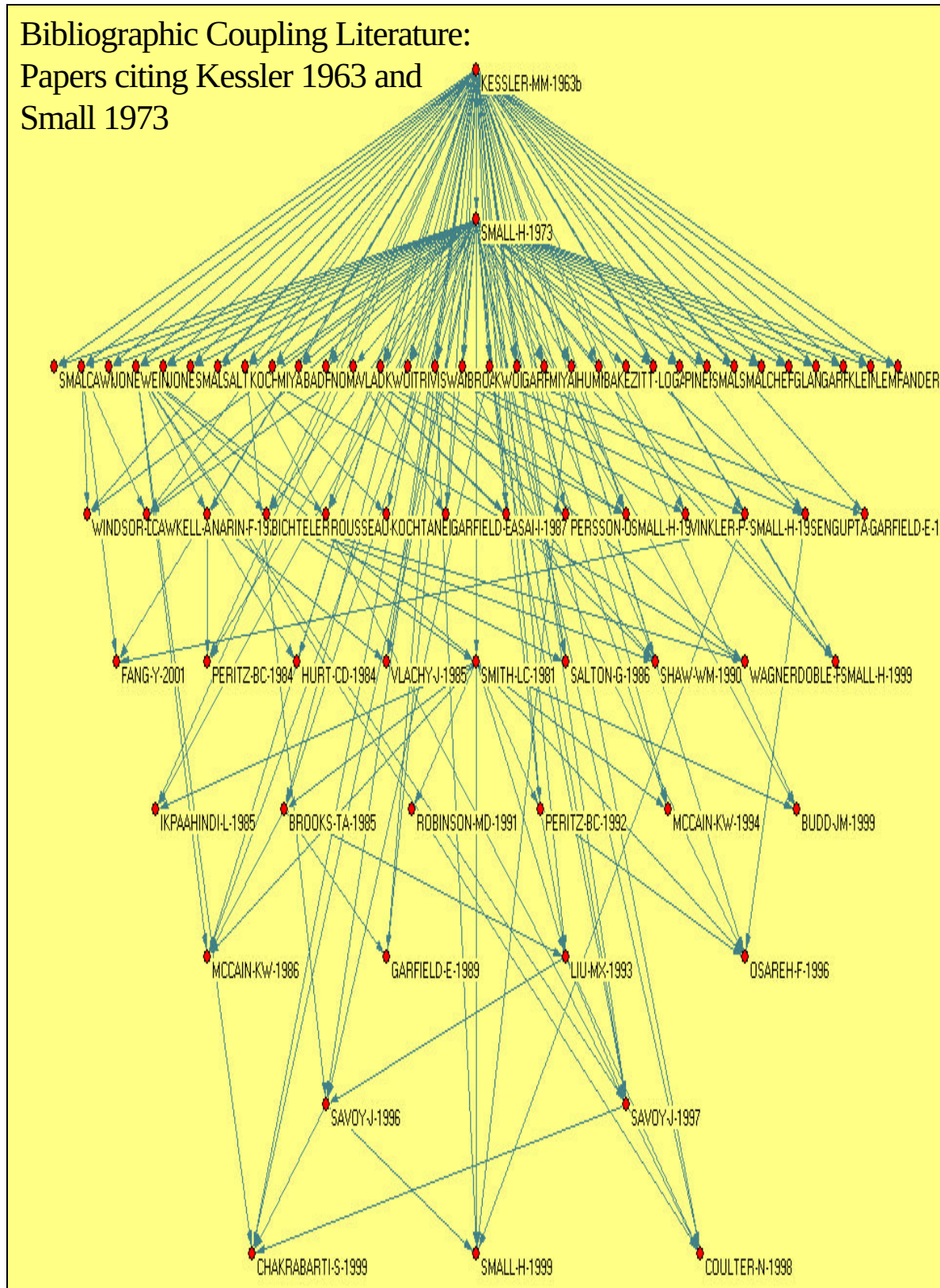
Main path analysis by Hummon / Doreian

Table 1
The 40 milestone events of DNA research from Garfield et al. (1964) after Asimov (1963)

Event	Date	Author(s)	Event	Date	Author(s)
1	1820	Bracconot	21	1947	Chargaff
2	1860s	Mendel	22	1950	Chargaff
3	1869	Miescher	23	1951	Pauling & Corey
4	1880	Flemming	24	1953	Sanger
5	1886	Kossel	25	1952	Hershey & Chase
6	1891	Fischer & Piloty	26	1953	Wilkins
7	1900	DeVries	27	1953	Watson & Crick
8	1900-10	Fischer	28	1953	DuVigneaud
9	1909	Levene	29	1955	Todd
10	1926	Muller	30	1953	Pallade
11	1928	Griffith	31	1955	Fraenkel-Conrat
12	1929	Levene	32	1955	Ochoa
13	1931	Alloway	33	1956	Kornberg
14	1935	Stanley	34	1957-8	Hoagland
15	1935	Levene	35	1960	Jacob & Monod
16	1936	Bawden & Pirie	36	1961	Hurwitz
17	1938	Caspersson & Schulz	37	1961	Dintzis
18	1941	Beadle & Talum	38	1961	Norvelli
19	1944	Martin & Synge	39	1962	Mirsky & Alibrey
20	1944	Avery et al.	40	1961	Nirenberg & Maltaei

There are many other routes to visualization of citation networks. Norman Hummon and Patrick Doreian demonstrated in 1989²⁰ how our 1964 data on the History of DNA could be used to create a critical path map (see page 52, figure 4) of this same topic.

SLIDE 13: BIBLIOGRAPHIC COUPLING HISTORIOGRAPH

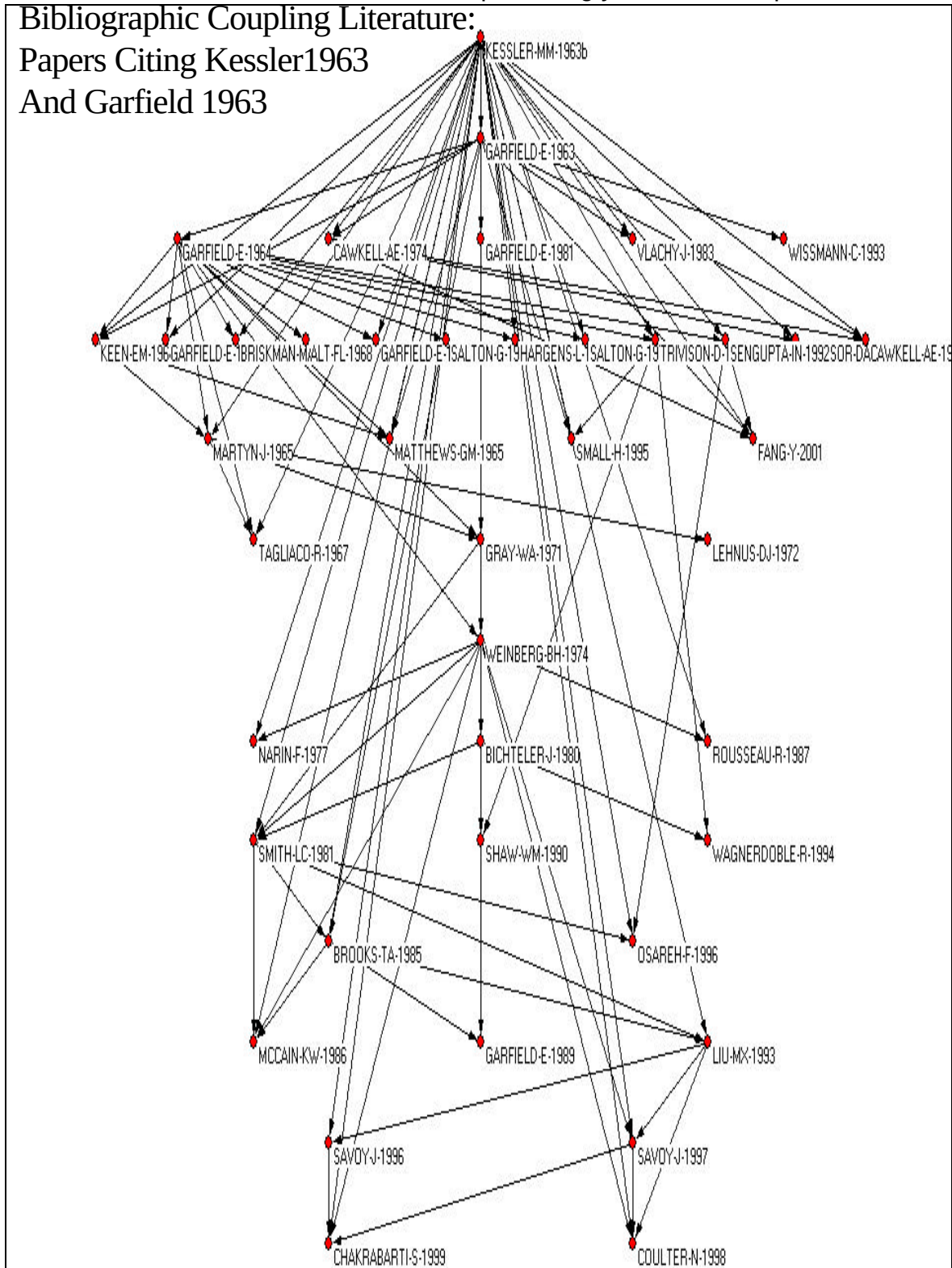


In future revisions, bibliographic coupling and co-citation methods may be used to group papers on extremely busy maps. These techniques should add additional dimensions to the visualization of evolving fields. Using the data in our bibliographic coupling file,

Howard White of Drexel was able to create co-citation based maps in Slide 13, based on Kessler and Small.

SLIDE 14: BIBLIOGRAPHIC COUPLING LITERATURE: PAPERS CITING KESSLER 1963 AND GARFIELD 1963

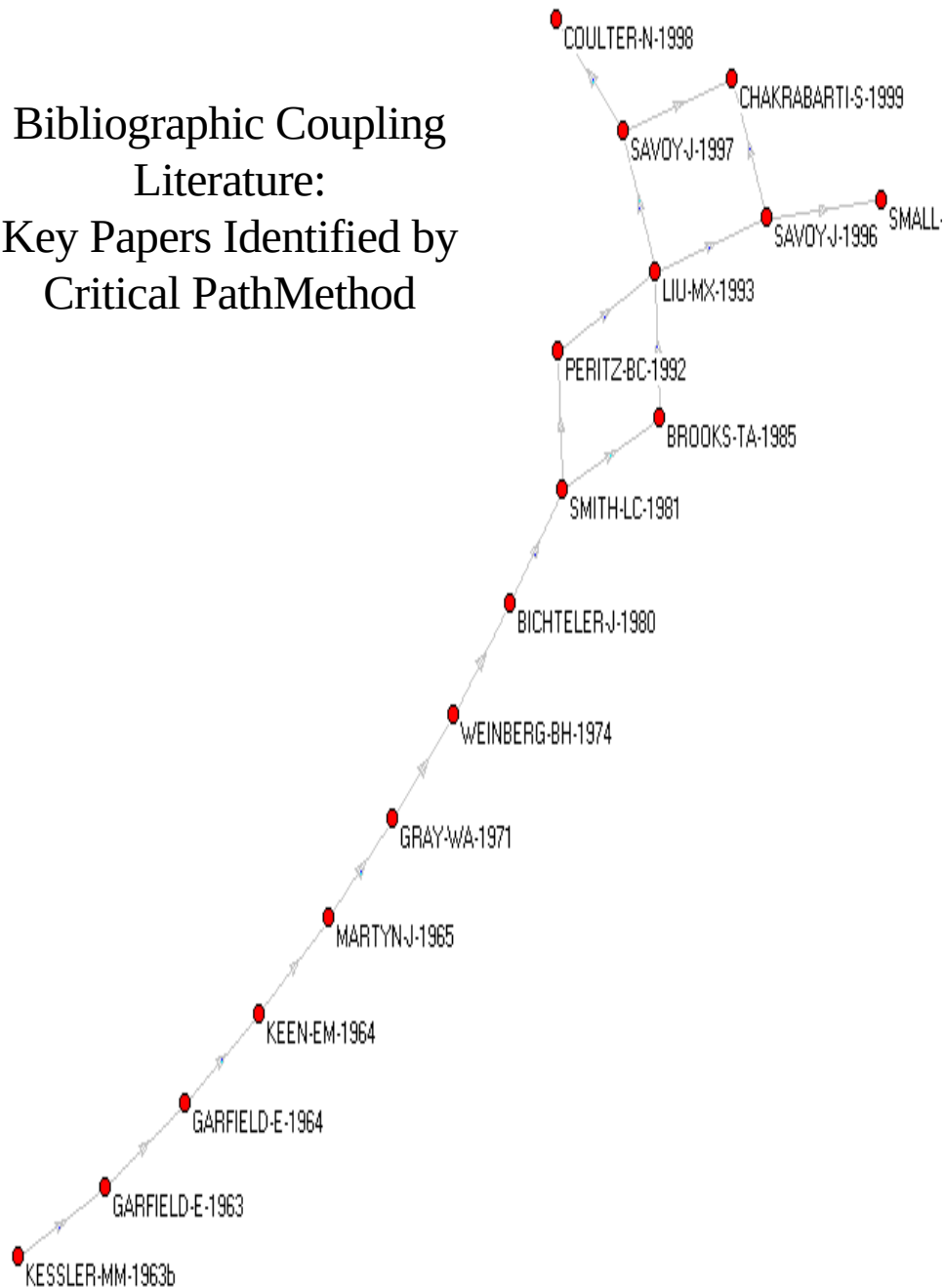
In Slide 14, he used Kessler and Garfield producing yet another map.



SLIDE 15: BIBLIOGRAPHIC COUPLING LITERATURE: KEY PAPERS IDENTIFIED BY CRITICAL PATH METHOD

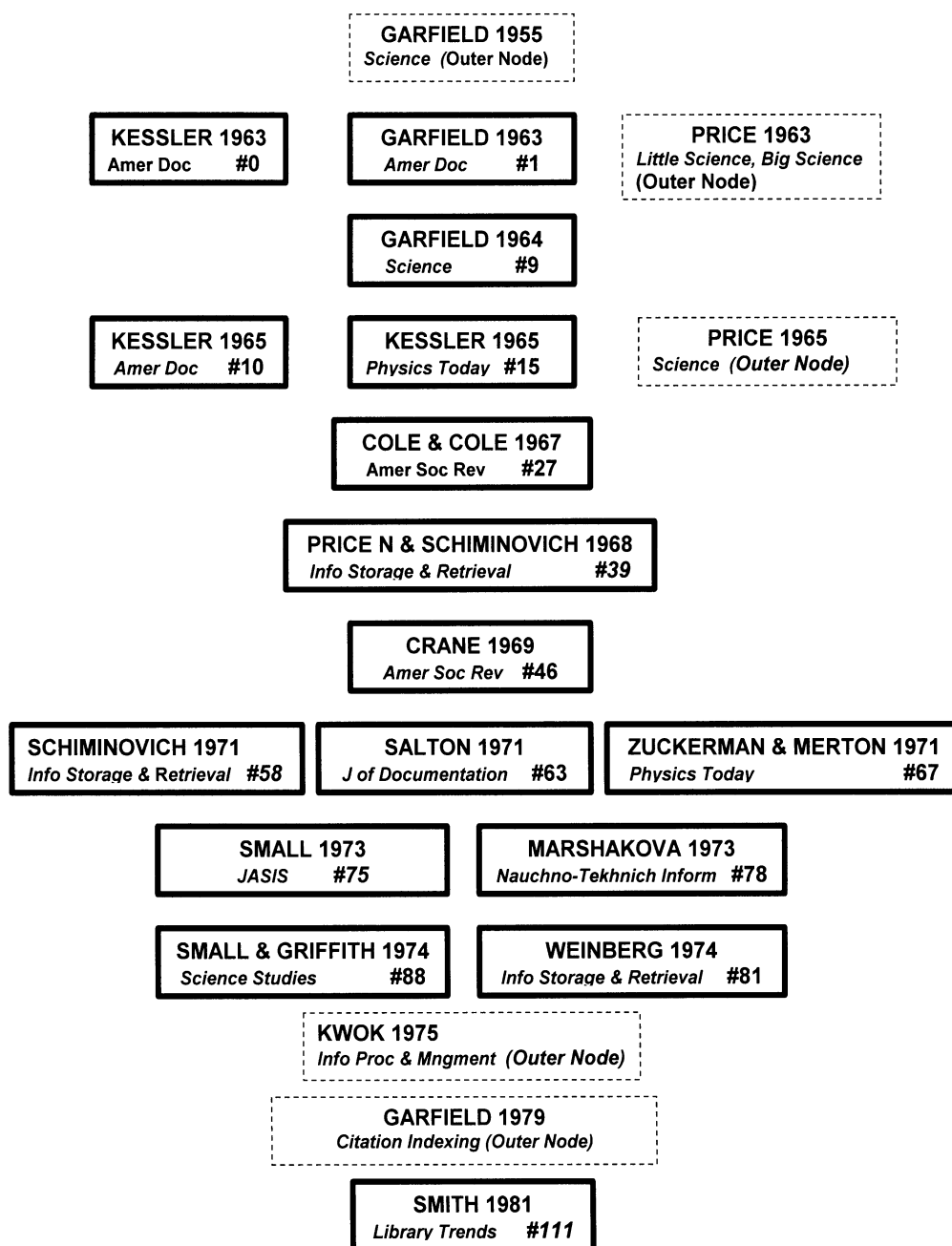
In Slide 15, using the Critical Path Method, he obtained another portrait of the evolution of this field. None of these gives a complete picture of the field which I have shown in the last Slide 16.

Bibliographic Coupling Literature: Key Papers Identified by Critical Path Method



SLIDE 16: FROM CITATION INDEXING TO BIOBLOGRAPHIC COUPLING TO CO-CITATION ANALYSIS

FROM CITATION INDEXING TO BIBLIOGRAPHIC COUPLING TO CO-CITATION ANALYSIS



It is remarkable that Kessler, eight years after my 1955 primordial paper on citation indexing, did not think it relevant to cite it in his 1963 paper. Nevertheless, a few months later, in *American*

Documentation and again in 1964 in *Science*, I cited him making it difficult for new readers to miss the connection.

It is interesting to note that Derek Price in his super-cited 1965 classic on "Networks of Scientific Papers" in *Science*²¹ did not cite the paper but referred to Kesssler's technical reports. In a footnote he acknowledged that he had used data from him as well as myself and others. You may have noticed in an earlier slide that Price's paper was the most-cited "outer node" making it quite relevant to the history of this topic. So were many other outer nodes like my 1979 book *Citation Indexing*²² and the report on the history of DNA.¹⁸ And that is the point of the exercise. Time does not permit me to go into further detail but the simple diagram I have created is a skeleton of a future map to be derived from these inputs.

Conclusion

I have traveled a long way with you from the days of my first meeting with Cas Borkowski on a hitchhiking tour. At one point, we even shared an apartment on 85th Street together. Through years of friendship, especially from 1949 to 1955 which I have often stated were the most important and productive years of my life, almost everything I later did at ISI could be traced back to those early years. My 1955 paper in *Science* culminated that early period.²³ Cas and I often talked about the potential benefits of creative bibliography to historians. So I am pleased to dedicate the first public preview of this software to him.

Thank you.

REFERENCES:

¹ Harris ZS, Hiz H, Joshi AK, Kaufman B, Chomsky E, and Gleitman L. *Transformations and Discourse Analysis Projects* (Department of Linguistics University of Pennsylvania, 1959-61).

² Harris ZS. "Linguistic Transformation for Information Retrieval." In: *Proceedings of the International Conference on Scientific Information, 1958*. Washington DC: National Academy of Sciences. Volume 2, pages 937-950 (1959)

³ Brownson, H. "Research on Handling Scientific Information," *Science* 132:1922-31 (1960).

⁴ Garfield, E. "Chemico-Linguistics: Computer Translation of Chemical Nomenclature," *Nature*, 192(4798):192 (1961)

<http://www.garfield.library.upenn.edu/essays/v6p489y1983.pdf>

⁵ Borkowski, CG. "Transformations of Polish." Philadelphia: University of Pennsylvania (1958).

⁶ Garfield, E. "From the World Brain to the Informatorium... with a little help from Manfred Kochen."

Presented at the University of Michigan, Ann Arbor, Symposium in honor of Manfred Kochen, September 21, 1999

http://garfield.library.upenn.edu/papers/kochen_worldbrain.html

⁷ Baxendale, P. "Vocabulary Control for Information Retrieval – F. W. Lancaster," *Library Quarterly* 44(3):256-258 (1974).

⁸ Palmquist R, "Class Lecture Notes: H.P. Luhn and Automatic Indexing – References to the Early Years of Automatic Indexing and Information Retrieval." (Spring 1998).

<http://fiat.gslis.utexas.edu/~ssoy/organizing/l391d2c.htm>

⁹ Bar-Hillel, Y. "The Present State of Research on Mechanical Translation," *American Documentation* 2(4):229-237 (1951)

¹⁰ Swanson, D. "Searching Natural Language Text by Computer," *Science* 132:1099-1104 (1960).

¹¹ Garfield, E. "Can criticism and documentation of research papers be automated? *Current Contents* No. 9 (March 4, 1970). Reprinted in *Essays of an Information Scientist*, Volume 1, pg. 83. Philadelphia: ISI Press (1977).

<http://www.garfield.library.upenn.edu/essays/V1p083y1962-73.pdf>

¹² Gafield, E. "Can Citation Indexing Be Automated?" in Mary Elizabeth Stevens, Vincent E. Giuliano, and Laurence B. Heilprin, Eds., *Statistical Association Methods for Mechanized Documentation, Symposium Proceedings, Washington, DC 1964* (National Bureau of Standards Miscellaneous Publication 269, December 15, 1965), pp. 189-192.

<http://www.garfield.library.upenn.edu/essays/V1p084y1962-73.pdf>

¹³ Lawrence, S. "Digital Libraries and Autonomous Citation Indexing," *Computer* 32(6):67 (1999)
<http://www.neci.nec.com/~lawrence/aci.html>

¹⁴ Small, HG. "Co-Citation Context Analysis – Relationship Between Bibliometric Structure and Knowledge," *Proceedings of the American Society for Information Science* 16:270-275 (1979)

¹⁵ Small H, Greenlee R. "Citation Context Analysis of a Co-Citation Cluster – Recombinant DNA," *Scientometrics* 2(4):277-301 (1980)

¹⁶ Small H. "Co-Citation Context Analysis and the Structure of Paradigms," *Journal of Documentation* 36(3):183-196 (1980)

¹⁷ Garfield, E. "Information theory and all that jazz: a lost reference list leads to a pragmatic assignment for students," *Current Contents* No. 44, pages 5-7 (October 31, 1977), Reprinted in *Essays of an Information Scientist*, Volume 3. Philadelphia: ISI Press, pp pgs. 271-273 (1980).

<http://www.garfield.library.upenn.edu/essays/v3p271y1977-78.pdf>

¹⁸ Garfield E, Sher IH, Torpie RJ. "The Use of Citation Data in Writing the History of Science." Philadelphia: The Institute for Scientific Information, 76 pgs. (December 1964)

<http://www.garfield.library.upenn.edu/papers/useofcitdatawritinghistofsci.pdf>

¹⁹ Asimov, I. *The Genetic Code*. New York: New American Library (1963).

²⁰ Hummon NP, Doreian P. "Connectivity in a Citation Network: The Development of DNA Theory," *Social Networks* 11(1):39-63, Figure 4 (1989)

²¹ Price, DJ. "Network of Scientific Papers," *Science* 149(3683):510+ (1965).

²² Garfield, E. *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*. Philadelphia: ISI Press, 274 pgs (1979).

²³ Garfield, E. "Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas," *Science*, 122(3159):108-11, July 1955.