

AMERICAN COMMITTEE  
for the WEIZMANN INSTITUTE OF SCIENCE

515 PARK AVENUE / NEW YORK, NEW YORK 10022 / PLaza 2-1300



STEPHEN L. STULMAN  
*President*

July 7, 1975

Rabbi Norman Lamm  
The Jewish Center  
131 West 86th Street  
New York, New York 10024

Dear Norman:

It has been too long since we saw you and Mindy. I trust all is well.

You are probably aware of the joint project between Bar-Ilan and Weizmann for Storage and Retrieval of The Responsa. I am sure that Dr. Fraenkel would like to have your comments.

Perhaps some of your congregants might be interested in becoming identified with this project through financial support. What do you think?

Warmest regards,

Sincerely,

A handwritten signature in blue ink, appearing to read 'Steve', is written over the typed name.

Stephen L. Stulman

SLS:mse  
enclosure

THE RESPONSA STORAGE AND RETRIEVAL SYSTEM

A project submitted by

Dr. Aviezri S. Fraenkel

of the

Weizmann Institute of Science.

### 1. Goal

The goal of the proposed research program is to make fundamental contributions to the problem of opening up the vast treasure-house of the Responsa literature, making its contents fully, easily, and effectively available to every researcher.

### 2. Background and Statement of Problem

The Responsa literature - written mainly in Hebrew and Aramaic, but containing vernaculars like Arabic, Ladino, Persian, German, Yiddish - is a collection of over half a million queries and answers spanning seventeen centuries. The queries were and are posed by individuals from all over the globe to the outstanding Jewish authorities in each generation. The decisions rendered became invaluable precedents for Jewish law, and many of them were subsequently incorporated into the codes.

The bulk of the problems dealt with are of a practical nature, reflecting real life situations. Thus the historical-sociological milieu is depicted in concrete episodes, in a true-life fashion usually unobtainable from other sources. Therefore the Responsa constitute an enormous store of material of interest to scholars in many areas, such as law, history, philosophy, religion, sociology, economics, linguistics, musicology, folklore, etc. Personages, realia, geographic sites, saints and scholars, wars and kings, together with those minutiae which bring the life of a community into sharp focus - birth, marriage and death customs, recipes, taxes, medical practices - all these

provide the scholar with a unique opportunity to recapture Jewish life with an accuracy often impossible from other sources.

The legal materials contained in the Responsa reflect in revealing detail the operations of the oldest applied legal tradition in the Western world. This system and philosophy of law have been applied under the most varied conditions and diverse countries, ranging from predominantly agricultural and rural societies to urban commercial and industrial states. It reflects the problems engendered over many centuries by changing economic, social and other conditions.

Part of this material is in manuscript, and the printed texts are scattered throughout the world, many unavailable even to the professional scholar. Moreover, there is no global index or information system encompassing any significant part of this literature. The Responsa project at Bar-Ilan University and The Weizmann Institute of Science was established to make a very fundamental contribution towards solving this problem.

### 3. Past Research Activities

To prepare the proper background for the presentation of the research program, we start with a description of past activities, which form the basis of the planned future activities.

The purpose of the project is to store the complete text of the Responsa in computers, and to develop original and sophisticated methods for the efficient and rapid retrieval of desired information from large masses of data.



The texts are first subjected to a short but thorough preediting procedure, such as incorporating footnotes into the text, and labeling of a few special passages or expressions. The complete text is then stored without any indexing, because otherwise readers would be bound by the information losses and distortions inherent in indexing, as well as by the limitations and restricted foresight of indexers [8].

For every new chunk of material stored, the computer generates several auxiliary files, which are merged with the corresponding files of previously stored material. Among them we mention the concordance and various dictionaries. The concordance gives for each text-word its exact coordinates, i.e., exact location within the sentences of a document. After this preparatory work, the texts are ready to be searched.

The searcher constructs a list of search terms characterizing his inquiry. Often, various precedents from the Talmud, the Codes, or earlier Responsa will have bearing on the topic. If any of these sources are known to the searcher, he can adjoin some of the pertinent languages used there to his search formulation. In fact, he can even adjoin the citation of the source as a compound search expression. The searcher also selects certain metrical operators, i.e., delimiters which relate the selected terms to one another in specified distance relations. Thus in effect he can specify expressions and phrases. All documents satisfying the specified conditions will be cited as relevant.

The above is somewhat of an oversimplification. In practice, each term selected by the searcher may appear in the text in several variations, which have to be adjoined to the search. If the texts would be in English, an alphabetical frequency list, giving for each word the number of times it appears in the text and the number of documents in which it appears, would solve the problem: It draws attention to the grammatical variants of the desired term, since usually these appear in the close vicinity of the selected word in an alphabetically ordered list. It also tells the searcher how many documents will be retrieved by use of a particular search term.

The highly inflected nature of Hebrew, however, and the omission of vowels in printed texts, imply that every word can have thousands of grammatical variants with different prefixes, infixes and suffixes - compared to less than ten in English. Moreover, these variants are scattered all about the alphabetically arranged frequency list, rendering it entirely useless.

To solve this problem, the computer was taught Hebrew grammar. In other words, algorithms were developed enabling the computer to conjugate any given search term into all legitimate grammatical variants, and printing out all those that actually appear in the texts stored in the computer. This print-out constitutes a local frequency list supporting the searcher for the search at hand. Moreover, a compact KWIC (Key-Word-In-Context) displaying the immediate vicinity in the texts of selected search terms can be scanned, giving valuable information about the search formulation. Thus, if a basic

source citation does not appear in the compact KWIC, the topic may not be dealt with seriously in the stored texts, signalling to the searcher to formulate a broad search, and/or select additional search terms. Conversely, if a given word appears very often, the context may reveal that it is homographic. In any case it cannot characterize the search, and additional conditions have to be imposed.

Clearly, the compact KWIC can also suggest relevant combinations of keywords, thus enabling the searcher to formulate a much improved and more precise search.

In the first stage of the project we developed the above described tools, experimented with them, refined and tested them [2, 3, 4, 6, 7, 9, 10, 11, 12, 16, 17, 18, 19]. After that we conducted an extensive test of the system by formulating 110 searches on a set of 2,473 Responsa, containing some 1,300,000 words [5]. A special team of manual searchers found close to 80% of the relevant material, whereas the computer cited 98% of the relevant material. At this time we stored about 18,000 Responsa comprising 13,000 words and conducted 250 in-depth searches for scholars all over the world.

During the Sixth World Congress of Jewish Studies in Jerusalem, sixty searches for Congress participants were conducted in three days, which engendered solutions to some research problems which researchers



7.  
have been working on for years. Some speakers at the Congress incorporated these new results in their talks. In the appendix we enclose the list of these sixty searches, as well as 26 searches conducted earlier this year.

#### 4. Future Research and Methods of Attack

(i) Self-modifying searches. A theory of self-modifying searches will be developed. In short, this means that partial results returned from the search are fed back to modify the original search formulation. To realize this, sophisticated mechanical grammatical tools are necessary. A first edition of a mechanized store of the entire Hebrew vocabulary was already created. New methods of local clustering aimed at adjoining additional relevant key-strings and rejecting non-relevant terms will make use of this store.

(ii) Storage of additional texts. In view of the success of the project hitherto, witnessed by the test results and the satisfaction of its various users, the interest in the project in wide circles of researchers is increasing. To service this growing community, tens of thousands of additional Responsa will be stored. A number of recognized authorities on the Responsa literature were asked to express their views on a policy of priorities of Responsa storage. Recently, the Rector of Bar-Ilan University appointed a committee of experts, which, together with members of the project team, formulated a policy of selecting books of Responsa to be stored. Specifically, a priority list of 100 books to be stored next was prepared.

(iii) The substantial increase in the amount of stored Hebrew texts will pose entirely new problems not encountered hitherto. This will require a fundamental review of past information storage methods, programs, and search techniques, the development of entirely new file

organization schemes and exploration of new directions of research, such as (i) above.

(iv) Hebrew concordance construction. The character of the Hebrew language, as pointed out before, renders the automatic construction of a Hebrew concordance a very difficult task. To minimize the manual editing involved, the concept of equivalence classes of forms depending on their neighbors was developed [13]. An extensive test of this approach will be conducted.

(v) Citation Index. Feasibility studies for citation index construction for the Responsa literature will be conducted. This research, conducted by request from abroad, is non-trivial, since some passages are cited without giving the source; in fact, sometimes without even acknowledging the passage as being a citation.

(vi) Computational linguistics. Various grammatical algorithms necessary for the above programs will be developed. Also, programs for syntactical analysis and homograph resolution will be developed.

(vii) Interactive communication. The system will be transformed into a conversational system accessed via terminals. This enables obtaining immediate answers to queries by rapid scanning of huge stores of texts.

The accomplishment of this program in its entirety is estimated to require a period of five years.



## 5. Significance

(i) The project does not produce an index for just another Responsa book. Rather, it attempts to solve the problem fundamentally and globally for all the Responsa. Moreover, whereas a conventional index constructed for a Responsa book is useful for a restricted class of users only - , e.g., index terms pertinent to law or history only are selected - our project addresses itself to all users, since the entire text is stored. For example, no system using a keyword index can enable linguistic research, in which, say, the occurrence of certain expressions is sought in order to trace the origin of a particular phrase.

(ii) The project is making available a vast treasure-house of information concerning all parts of the civilized world during the past 1700 years. This material is of great importance in various areas of the humanities, including linguistics, the development of language, history, philosophy, jurisprudence, and anthropology. Thus, for example, the materials contain detailed information on the history of various communities, the social, religious, economic and other problems that confronted them at different times and under a variety of circumstances, the solutions adopted and their effect, as well as the relationship between various religious and geographical groups over long periods of time.

(iii) It is making available material concerning the development of languages over a long time span, under the influence of various cultural, religious, social and other factors. The Responsa are not only written in various languages, particularly Hebrew, Aramaic and Arabic, over many centuries, but also contain a large number of references to terms used in other languages and to a variety of geographic locations.

(iv) The project will contribute greatly to computational linguistics. The development of efficient information retrieval systems requires detailed linguistic analysis and grammatical synthesis in order to search out all materials relevant to any required topic, regardless of the different forms, phrases, and words utilized in the materials. Techniques have been developed by the project and are being perfected, which are of great importance to linguistics, such as grammatical synthesis, including the formulation of algorithms to derive automatically all grammatical variants of any given term, and to indicate to the researcher which grammatical variants are actually present in the materials being searched. The project has developed and continues to perfect the automatic generation of concordances and dictionaries.

(v) It is developing powerful tools for research in the humanities through the perfection of techniques for automatic storage and retrieval of information without indexing.

This will permit storage and rapid efficient retrieval of all kinds of information in the humanities. The techniques perfected for the project will save vast amounts of storage space in libraries and other repositories of knowledge and enable much faster and more efficient search of information in the humanities and in all of the social sciences. These techniques are rapidly becoming more vital, indeed indispensable, because the rapidly accelerating outputs and accumulation of data are quickly becoming unmanageable without the use of computers.

(vi) The project contributes significantly to legal research, especially in case law. It is recognized that mechanized storage and retrieval of case law is more difficult than statutory law, because judicial decisions are written more loosely and with a wider range of language than statutes. In addition, research in case law is often aimed at basic rationales of a decision and at legal concepts contained therein. This makes it more difficult to track down by use of specific search words, than statutory materials which tend to be more precisely formulated and which repeatedly utilize the same phrases. The techniques developed by the project for research in the Responsa, which consist in large part of Hebrew case law, will easily be applicable for legal research which involves less complexity of language. It will therefore have tremendous impact on legal research in general.



(vii) It is creating powerful tools in the theory of automatic storage and retrieval, such as the concept of self-modifying searches, local clustering based on defining eigenvectors of terms, new file organization methods, computational linguistics, and the imbedding of the concept of a citation index in the full text method. Dr. Eugene Garfield, President of the Institute of Scientific Information, in writing about the project in [15], pointed out especially the last two aspects of the project as being of special interest.

(viii) The activities of the project attract a number of bright graduate students in computer science, linguistics and other disciplines. Other graduate students, mainly in the humanities and law, found out that the project can greatly help them in their own thesis work, by retrieving material in minutes that they had been looking for for years.

#### 6. References

1. R. Attar, Y. Choueka, A. S. Fraenkel and D. Schindler, Linguistic tools and local clustering processes in information systems. In preparation.
2. Y. Choueka, M. Cohen, J. Dueck, A. S. Fraenkel and M. Slae, Full text document retrieval: Hebrew legal texts (Report on the first phase of the Responsa retrieval project), Proc. Symp. on Information Storage and Retrieval (J. Minker and S. Rosenfeld, eds.), University of Maryland, pp. 61-79, April 1-2, 1971.
3. Y. Choueka, M. Cohen, J. Dueck, A. S. Fraenkel and M. Slae, Full text case law retrieval: the Responsa project (a solicited expanded form of paper no. 2), Working Papers on Legal Information Processing Series, Pamphlet No. 3 (64 pages), J. Schweitzer Verlag, 1972.
4. Y. Choueka, A. S. Fraenkel and M. Slae, Full text case law retrieval: the 2<sup>nd</sup> phase of the Responsa project. In preparation.
5. Y. Choueka, A. S. Fraenkel, Y. Pechenek and M. Slae, The Responsa retrieval project: methods, tests, results. In preparation.
6. M. Cohen, Full text information retrieval (in Hebrew), M. Sc. Thesis submitted to Tel-Aviv University, 1970.
7. A. S. Fraenkel, Seventeen centuries on tape, Law and Computer Technology 1, No. 2, pp. 8-9, February 1968.

8. A. S. Fraenkel, Legal information retrieval, in *Advances in Computers* (F. L. Alt and M. Rubinoff, eds.) 9, pp. 113-178, Academic Press, 1968.
9. A. S. Fraenkel, Modern retrieval methods applied to Jewish case law, *Proc. First Intern. Symposium Soc. Tech. Writers and Publishers*, Tel Aviv, pp. R1-1 — R1-9, December 9-12, 1968.
10. A. S. Fraenkel, A retrieval system for the responsa, *Proc. AOJS* 2, pp. 3-42, 1969.
11. A. S. Fraenkel, Full text document retrieval, *Proc. Symp. on Information in Chemistry*, Tel Aviv, pp. 86-90, May 14-15, 1969.
12. A. S. Fraenkel, Mechanized information retrieval as an auxiliary tool in Jewish law research, *Proc. Fifth World Congress of Jewish Studies*, Jeruslaem, 1969 (abstract in English; full article in Hebrew), vol. 5, Jerusalem, pp. 79-98, 1973.
13. A. S. Fraenkel and E. Spitz, The theory of Hebrew concordance construction. In preparation. (Abstract and lecture at Sixth World Congress of Jewish Studies in Jerusalem, 1973.)
14. A. S. Fraenkel and J. Stein, Local feedback methods in information retrieval systems. In preparation.
15. E. Garfield, Full text searching systems just around the corner?, *Current Contents (Life Sciences)*, No. 36, pp. 7-8, Sept. 1973.
16. A. M. Schreiber, Computerized storage and retrieval of case law without indexing: the Hebrew Responsa project, *Law and Computer Technology*, pp. 14-22, November 1969.
17. M. Slae, Marranos' search in the Rivash responsa, *Bar-Ilan Yearbook*, Vols. 7-8, Hebrew section: pp. 397-419; English section: pp. XLVII-LXIII, 1969-1970.
18. M. Slae, The occult in the Rivash responsa (Hebrew), *Shanah B'Shanah* 5731, pp. 226-236, Heichal Shlomo, Jerusalem, 1970.
19. M. Slae, Legal research through computers: the Responsa project (Hebrew), *Dine Israel Vol 4*, Faculty of Law, Tel-Aviv University, pp. 233-245, 1973.



# Curriculum Vitae

AVIEZRI S. FRAENKEL

## 1. Personal.

Born: June 7, 1929, Munich, Germany  
Arrival in Israel: February, 1939 (from Switzerland)  
Status: Married (1956), six children  
Military service: December 1947 - Summer 1950, and  
November 1951 - October 1954

## 2. Education

|      |  |
|------|--|
| 1953 | B.Sc. and Diploma in Electrical Engineering<br>Technion-Israel Institute of Technology       |
| 1957 | M.Sc. (with distinction), Electrical Engineering,<br>Technion-Israel Institute of Technology |
| 1961 | Ph.D., Mathematics, U.C.L.A.   |

## 3. Experience

|                         |   |
|-------------------------|---|
| 1953 - 1954             | Electronic engineering, Israel Army   |
| Oct. 1954 - Summer 1967 | Member of computer group, Weizmann Institute<br>Participated in design and construction of<br>WEIZAC, the first Electronic Digital Computer<br>in Israel.     |
| Fall 1957 - Summer 1961 | Student and Research Assistant, Department<br>of Engineering, U.C.L.A. (Fulbright Travel<br>Grant). Changed major from Engineering to<br>Mathematics in 1959. |
| Academic year 1961/62   | Assistant Professor, Department of Mathematics,<br>University of Oregon.  |
| 1962 - present          | Senior Scientist, Department of Applied<br>Mathematics, Weizmann Institute.   |
| 1964 - 1967             | Member, Executive Committee of IPA - Information<br>Processing Association of Israel.   |
| 1965 - 1966             | Chairman IPA High-School Computer Education<br>and Curriculum Committee.  |
| Academic year 1966/67   | Associate Professor and Chairman, Department<br>of Mathematics, Bar-Ilan University (leave<br>of absence from Weizmann Institute).                            |
| 1967 - present          | Associate Professor, Department of Mathematics,<br>Bar-Ilan University (on part-time<br>basis).   |

|                |   |
|----------------|---|
| 1967 - 1969    | Chairman, Department of Mathematics, Bar-Ilan University. Head, Computing Center Development Project, Bar-Ilan University (on part-time basis). |
| 1968 - 1970    | Member, IFIP TC-1 Committee on Secondary School Education/Computer Science.   |
| 1968           | Article on Computers - Technological Principles of, in Israel Encyclopedia.   |
| 1969           | Recipient, Royal Society Award, for attending "Computers in Number Theory" Symposium in Oxford, England, 1969.                                  |
| 1969 - 1970    | Sabbatical leave at Algebra and Number Theory Year, Department of Mathematics, University of California, Los Angeles.                           |
| Sept. 1970     | Invited participant at National Bureau of Standards Workshop on Information Storage and Retrieval, Washington, D. C.                            |
| 1970 - 1972    | Chairman, Israel Mathematical Union.  |
| 1971           | Recipient, NSF grant to attend "Combinatorial Theory Advanced Science Seminar" at Bowdoin College, Brunswick, Maine.                            |
| 1971 - 1974    | Member, Executive Committee of IPA.   |
| 1971 - 1972    | Associate Professor, Department of Mathematics, Tel-Aviv University (part-time).  |
| 1972           | Recipient, Feder Foundation Prize of the AOJS.  |
| 1972 - present | Member, Editorial Board, Datenverarbeitung im Recht, Schweitzer Verlag.   |
| 1972 - present | Member, Israel Committee for Codata.  |
| 1972 - present | Editor, The Bulletin of IPA.  |

## GRANTS

1. Computational Techniques and Number Theory - National Bureau of Standards, 1965-1967.
2. Computational Techniques and Number Theory - National Bureau of Standards, 1968-1969.
3. Full Text Retrieval, Application to Classification and Other Uses - National Bureau of Standards, 1971-1973.
4. Responsa Retrieval Project - National Endowment for the Humanities, 1971-1975 (Associate Investigator).
5. Royal Society, summer 1969: Travel and subsistence.
6. National Science Foundation, 1969-1970: Partial support to participate at Algebra and Number Theory Year, University of California, Los Angeles.

7. National Science Foundation, Summer 1971: Subsistence.
8. National Research Council of Canada, summer 1972: Partial support while visiting the University of Calgary.
9. National Science Foundation, summer 1973: Partial support to participate at AMS Conference on Influence of Computers on Mathematics, Missoula, Montana.

Who's Who in Israel, Hebrew Edition (1973), English Edition (1973/74).

Dictionary of International Biography XI (England, 1974).