

December 17, 1964

Dr. C. Salton
Computation Laboratory
Harvard University
Cambridge 38, Massachusetts

Dear Dr. Salton:

I particularly want to thank you for sending me the report from your group on Computational Techniques in Information Retrieval, and Analysis. The work that you and Dr. Sussenguth have been doing on the retrieval of chemical structures should be of particular interest to me, though I must confess I have not yet expended the effort that I would need to reach a full understanding of your mathematical approaches.

I am enclosing a copy from proof of a paper that will appear soon in the Proceedings of the National Academy of Sciences. This gives a rather bare outline of an approach that I have been developing, more for hypothesis formation than for information retrieval, though it is obvious that there is a good deal of convergence between these objectives. There is, as always, an enormous amount of detail involved in a complete documentation and I have been busy trying to write up the full report and will gladly send you copies as soon as this is ready. The descriptions in the enclosure are not necessarily in the final canonical form, although the general ideas have not been basically altered.

As far as I could tell there was no great redundancy between these suggestions and what your group has been doing; instead it is just possible that these efforts may be able to complement one another, I hope in some useful way. In particular, I would now suppress any vagrant thoughts that I might otherwise have had about techniques of searching for substructures, at least until I could indeed fully understand the techniques that are outlined in your reports. On the other hand, I would like to offer the thought that for most real organic molecules it is possible to prescribe a canonical form without inordinate difficulty. This may well be a useful step in the kind of analysis you have been doing as it has been for my own.

Sincerely yours,

Joshua Lederberg
Professor of Genetics

SALTON ? G.