Essays of an Information Scientist, Vol:1, p.302-303, 1962-73 Current Contents, #19, May 17, 1972 Also See: Vo1p304-1962-73.pdf

"Current Comments"

ISI's *Chemical Substructure Index* -- Research Resource and Invaluable Tool for Health and Environmental Sciences!

Click for Reprint: Charles Granito & Murray Rosenberg "Chemical Substructure Index (CSI) A New Research Tool, p.304, Essays, Vol:1

May 17, 1972

The vast billion dollar programs envisioned to conquer cancer and other diseases will require not only a stockpile of usable chemical information, but also an available system for handling information from the new research.

In this issue of *Current Contents®*, we are reprinting a recent article[1] on ISI®'s *Chemical Substructure Index™ (CSI™)*.[2] The *CSI* appears monthly and is cumulated annually. *CSI* enables you to locate chemical compounds by broad structural type or particular moiety or substructure. In addition to its obvious applications to cancer research, *CSI* is pertinent to our nation's commitment to solve problems of pollution, drug abuse, and other health related problems too numerous to mention.

The *CSI* uses the Wiswesser Line Notation[3] to encode structures of all new compounds reported in articles abstracted in ISI's *Current Abstracts of Chemistry and Index Chemicus™*.[4] Each notation describes every atom and bond in a three-dimensional chemical structure, and each notation is then indexed under every significant structural symbol used in describing the compound.

This "permutation" or rotation of each Wiswesser notation, and its multiple entry in the alphabetic listing of substructure symbols, assures that the user will find relevant compounds no matter from what particular structural viewpoint he may consider them to be of interest. If you should, for example, be interested in iminopropynylthiazolines, you'll find them, whether you look them up as imino compounds, propynyl derivatives, or thiazolines.

Use of the Wiswesser notation with a multiple indexing of every encoded compound in its permuted variations offers tremendous advantages over conventional alphabetic indexes based on traditional chemical nomenclature. In the first place, no traditional nomenclature, such as *Chemical Abstracts,* has an indexable term to describe all kinds of structural combinations a user may be seeking. No traditional nomenclature can deal with "heterocyclic carbons doubly bonded to any substituent atom except oxygen", or "the sequence $-CH_2-CH=N-$ whether the carbons are cyclic or acyclic". Even if a usable word existed to express such concepts as index terms, traditional nomenclature is hierarchical. Some features must be subordinate to others, the order determined by complicated rules of priority and position. As a result, compounds similar in some respects but different in others must of necessity be scattered throughout the index. In the *Chemical Substructure Index* all compounds with similar structural features are grouped together, however different the compounds may be in

other respects. This makes it possible to do desk-top searches that some people imagine require a computer. Indeed, until *CSI* came along this multi-dimensional searching capability did require a computer.

Chemists who have tried mastering any of the traditional nomenclatures may tend to shy away from a system of chemical notation, but they're wrong to do so. In general, the Wiswesser notation is much easier to learn than any system of nomenclature. What the chemist needs to know in order to use the *CSI* can be learned in half an hour. This is true partly because the notation achieves the aim of any code as a communications medium: it is systematic, it is basically simple, and the notations are admirably brief. But it is primarily true that the user can become proficient with the *CSI* in a short time because, as with any "language", it is much simplier to learn to read (decode) than to write (encode). The article reprinted in this issue makes clear how easy the whole business is.

The *CSI* is a record of part of each month's input to ISI's chemical data bank, which we have been building for more than ten years.[5] The data bank is accessible by a variety of means: magnetic tape, printed abstract and index volumes, and an SDI automatic alert system. It is, thus, an invaluable resource as well as an invaluable tool for research. ISI's chemical data bank and chemical data system qualify uniquely in both respects.

1. Granito, C.E. & Rosenberg, M.D. *Chemical Substructure Index™ (CSI ™)*, a new research tool. *J. Chem. Doc.* 11(4):251-256, 1971.
2. Garfield, E. New *Chemical Substructure Index* is creative theoretical tool for molecule manipulators as well as practical system for retrieval. *Current Contents* No. 24, p. 5-6, June 16, 1971.
3. Gibson, G.W. & Granito, C.E. Wiswesser chemical line-notation. *American Laboratory,* April 1972, p. 27-37.
4. Garfield, E. *Current Abstracts of Chemistry and Index Chemicus Current Contents* No. 49, p. 4-5, December 2, 1969. Reprinted in: *Chem. Eng. News* 47(51):66, 1969, and *Chem. Industry,* January 10, 1970, p. ii.
5. Garfield, E. et al. *Index Chemicus Registry System* , pragmatic approach to substructure chemical retrieval. *J. Chem. Doc.* 10:54-58, 1970.