

CAC/IC Strikes Again!  
A Computer-Output-Microform (COM)  
Index to 1.25 Million New Compounds

April 7, 1975

Number 14

Recently, ISI® made available an index to more than 1.25 million new compounds that were reported in leading chemical journals from 1966 to 1974. These 'new' compounds include not only newly reported compounds, but also old compounds in newly reported syntheses and reactions.

This new compilation is available on only seventy-four microfiches, which fit into a four-by-six inch box one-quarter inch thick. If you're inclined to spend \$1000, the new index will fit into six cubic inches of your vault or desk, whichever is more convenient.

The 1.25+ million compounds covered—and recoverable—by the new cumulative molecular-formula index were processed for 1966–1974 issues of ISI's *Current Abstracts of Chemistry and Index Chemicus™ (CAC/IC™)*. That's a considerable and important mass of chemical information that you can now have literally at—and between—your finger tips.

The new microfiche index is arranged according to the traditional molecular formula system. However, after each molecular formula the Wiswesser Line Notation (WLN) for each compound is provided. A separate WLN is given for homologous compounds with the same molecular formula. Each WLN 'sub-entry' refers the user to a unique registry number. This number identifies the abstract in *CAC/IC*, and the specific structural diagram in that abstract.

Each of the seventy-four fiches in the index has 202 frames. Two of the frames on each fiche index the material contained on the particular fiche—that is, they indicate to which frame you should go for a

particular range of molecular formulas. Each fiche is itself labeled with its full formula range across the top.

I know of no other molecular formula index that offers in such small and easily usable form access to such a mass of the chemical literature reporting new compounds. There are 288,439 unique molecular formula entries on these fiches. They cover 1,227,503 unique compounds, each with a WLN, or about 4.25 compounds per molecular formula.

In compiling the index, we discovered something interesting and to a certain extent disturbing. Of the 1,319,742 WLN's, 92,239 were duplicates. That figure is just about 8% of the true unique-compound total. There is one and only one correct WLN for any compound. And no two compounds can have the same WLN. The notations, in encoding jargon, are both unique and unambiguous. The duplicate notations mean that too frequently a compound reported or implied to be new hasn't been. It's not surprising that that should happen once in a while. But 8% is much too high a figure for 'chance' duplication.

In compiling the microfiche index, we found that the duplications were sometimes the work of different researchers working independently—but not often. In most cases, it was the same author or group publishing in different journals without referencing slightly earlier work. Were they trying to get more bang for the buck? It's an unfortunate practice, and is the subject of an ISI paper that will be presented at the forthcoming meeting of the American Chemical Society in Philadelphia this month.<sup>1</sup>

Readers may want to become familiar with the acronym *COM* —computer-output-microform. I believe that microforms will become increasingly used as their preparation is automated by this method. Our new index is *COM*. The method reduces the cost and time of hard-copy preparation from data banks with huge amounts of digitalized information. In the old days when it was necessary to 'dump' the computer files, you had to print them out. The printout was then photographed for off-set printing. This was a long and expensive process. In the new method, the computer information is printed directly onto film. This not only speeds up the output by at least an order of magnitude, but also eliminates the printing process.

We also plan to encode another half-million compounds indexed by *CAC/IC* during the years 1960 through 1965. At that time, the present 1966-1974 file will be completely updated to include not only the 1960-1965 data, but also compounds reported in 1975.

I've written here before about *CAC/IC* and about *WLN*.<sup>2,9</sup> Readers still unfamiliar with either can write to me for more information. Or you might prefer to complain that we've neglected you.<sup>10</sup> Copies of the referenced materials on *CAC/IC* and *WLN* are available on request.

The new microfiche index to *CAC/IC 1966-1974* is ready for delivery now. If all of this interests you, please let me know.

1. Batzig J H & Lawlor H A. "The ethical impact of 'publish or perish' on the increase of new compounds reported in the chemical literature. I." Paper to be presented at the National Meeting of the American Chemical Society, Philadelphia, 10 April 1974.
2. Garfield E. The 200th anniversary of *Index Chemicus*. *Current Contents (CC)* No. 44, 10 November 1967, p. 4.
3. -----, *Current Abstracts of Chemistry and Index Chemicus*. *CC* No. 48, 2 December 1969, p. 5-6.
4. -----, The retrieval and dissemination of chemical information. *CC* No. 28, 15 July 1970, p. M1-2.
5. -----, The retrieval and dissemination of chemical information. II. The Wiswesser Line Notation. *CC* No. 29, 22 July 1970, p. M1-2.
6. -----, *New Chemical Substructure Index* is creative theoretical tool for molecule manipulators as well as practical system for retrieval. *CC* No. 24, 16 June 1971, p. 5-6.
7. -----, ISI's *Chemical Substructure Index* — research resource and invaluable tool for health and environmental sciences. *CC* No. 20, 17 May 1972, p. 3-4.
8. -----, Are you ready for chemical linguistics? chemical semantics? chemical semiotics? Or, why *WLN*? *CC* No. 50, 13 December 1972, p. 5-7.
9. -----, We've added a Weekly Substructure Index to *Current Abstracts of Chemistry and Index Chemicus*. *CC* No. 5, 3 February 1975, p.5-6.
10. -----, Don't kill us with kindness—*COMPLAIN!* *CC* No. 6, 10 February 1975, p.5-6.