On Randić's Molecular Identification Numbers

K. SZYMANSKI, W. R. MÜLLER, J. V. KNOP,* and N. TRINAJSTIĆ†

Computing Centre, The University of Düsseldorf, 4000 Düsseldorf, Federal Republic of Germany

Received February 8, 1985

The search for counterexamples to Randić's molecular ID numbers in the field of alkane trees up to 20 atoms produced 124 duplicates and 1 triplicate. Thus, it is shown that the ID numbers, although highly discriminating indices, are not unique.

Randić has recently introduced in this journal a new topological (or more correctly, graph theoretical) index, named molecular identification (ID) number, in order to identify graphs by a real number.[1] He defined ID number in the following way. Let G be a graph with the vertex set V and the edge set E. If $v_1, v_2 \in V$, we denote by $e = (v_1, v_2) \in E$ the edge connecting v_1 to v_2. By deg v we denote the degree of a vertex v in G, i.e., the number of edges incident with v.[2]

Randić defined the mapping from the edge set to the real numbers

$$g: E \rightarrow \mathcal{R} \tag{1}$$

by

$$g(e) = g(v_1, v_2) = \frac{1}{(\deg v_1 \times \deg v_2)^{1/2}} \tag{2}$$

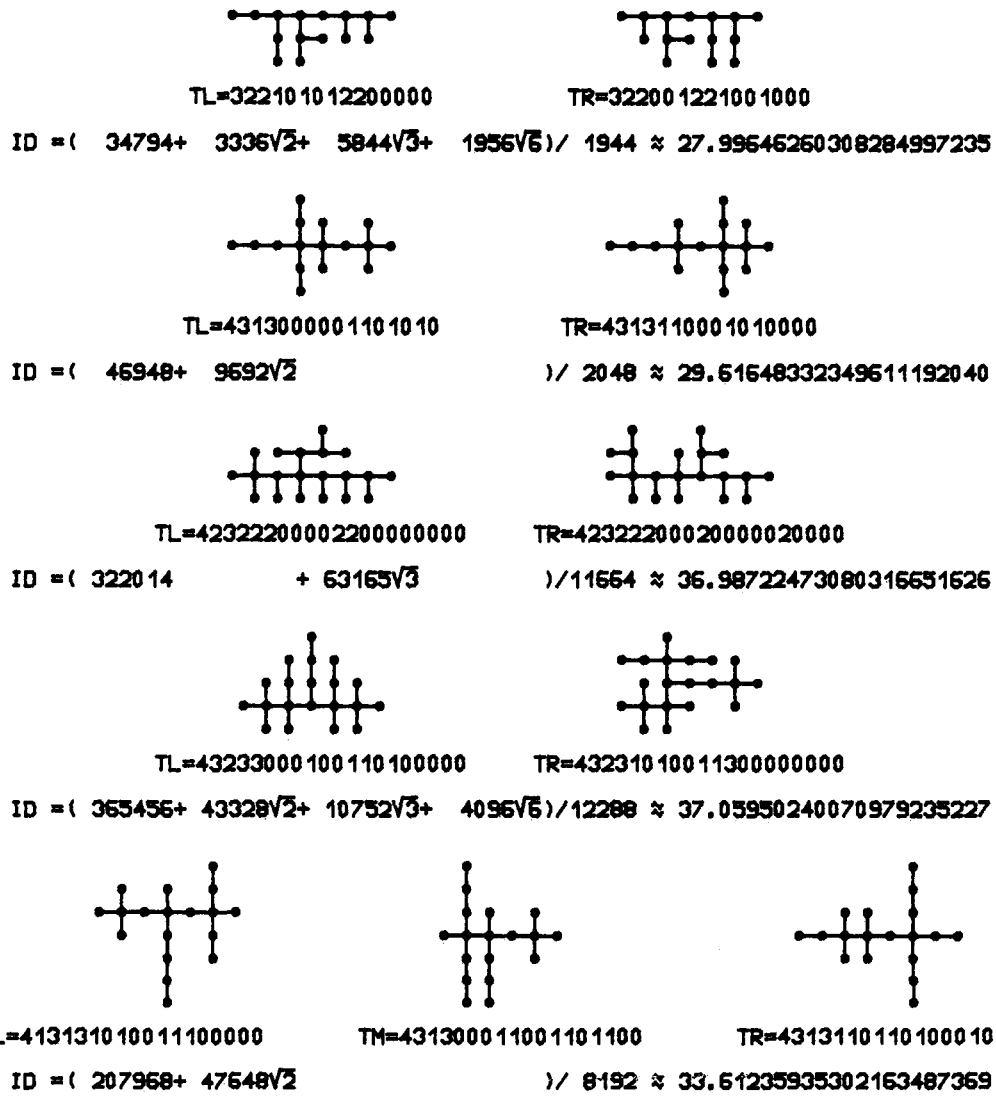

Figure 1. Some alkane trees with exactly equal ID indices. TL, TM, and TR stand for left tree, middle tree, and right tree. The first set of numbers beneath each tree is the N-tuple representation of the tree (see reference 4).

† Permanent address: The Rugjer Bošković Institute, 41001 Zagreb, Croatia, Yugoslavia.

0095-2338/85/1625-0413$01.50/0

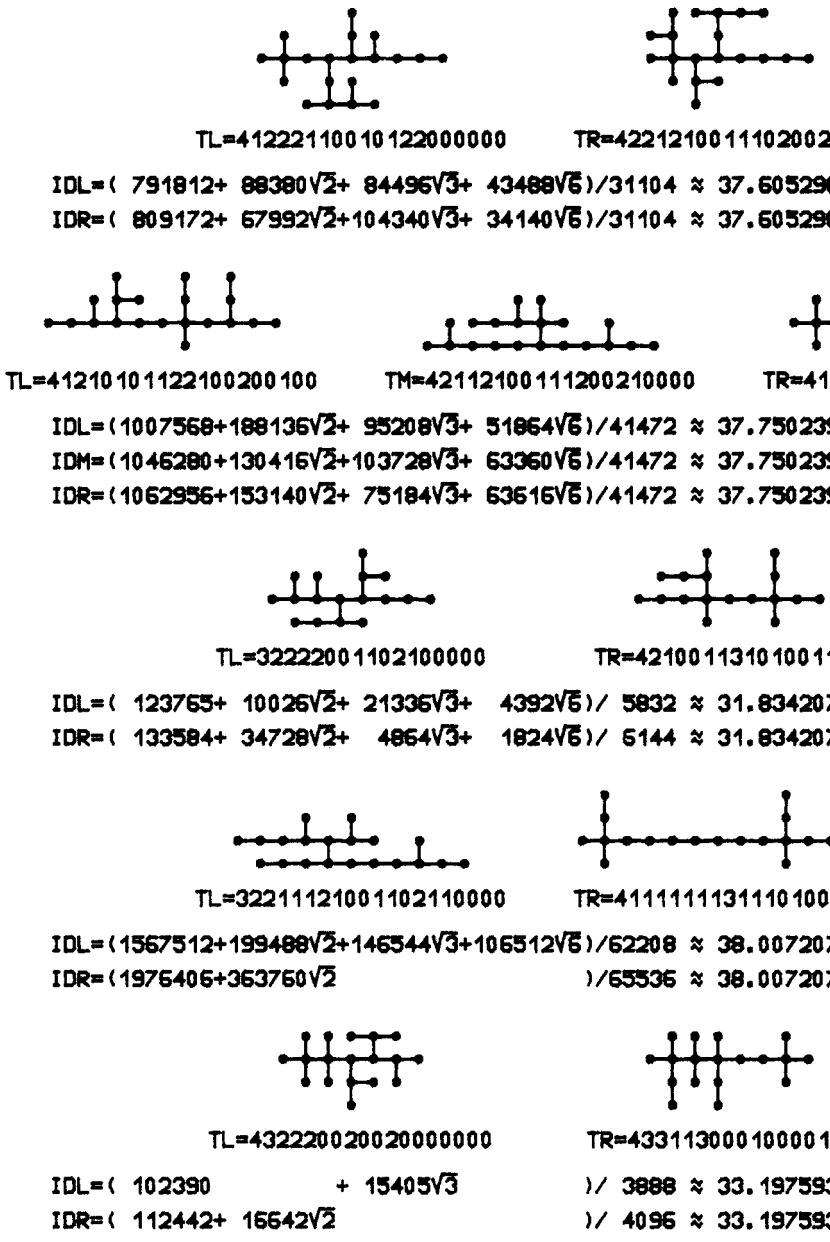

Figure 2. Some alkane trees with nearly equal ID indices. TL, TM, and TR stand for left tree, middle tree, and right tree. The first set of numbers beneath each tree is the N-tuple representation of the tree (see reference 4).

Let $p = e_1, ..., e_m$ ($m > 0$) be a path in G; then, the mapping g can be extended to the set of paths in G by

$$g^*(p) = \prod_{i=1}^{m} g(e_i) \tag{3}$$

The ID number of G is then finally defined as

$$\mathrm{ID}(G) = N + \sum_{p} g^*(p) \tag{4}$$

where N is the number of vertices in G and the summation is taken over all different paths in G.

The ID number appears to be an attractive topological index that is relatively easy to derive and has structural significance. The question: Are the ID numbers unique?, however, was not answered by Randić. He examined over 400 graphs and found no pairs of graphs with the same ID number. In this paper we will try to answer on the above question for alkane trees. Alkane trees are trees representing alkanes.[2] A tree is a connected acyclic graph. A graph is acyclic if it has no cycles.[3]

Since we are in position to generate all molecular trees (and subsets of trees) up to a given number of vertices,[4-6] we computed the ID numbers for all alkane trees up to 20 vertices (atoms). The following result was obtained: In the field of 618 050 alkanes (all alkanes up to 20 carbon atoms) there are 1975 pairs and 10 triples of nonisomorphic structures having the same ID numbers. This computation has been done with a single precision (by a word length of 48 bits). In order to check on this result and to find the most accurate value of the ID numbers for this class of alkanes, we used our own arithmetic (computing in terms of 1, $2^{1/2}$, and $3^{1/2}$ over the field of rational numbers),[7] which uses only integers in computer operations. Now, we found only 124 pairs and 1 triple of nonisomorphic alkanes having exactly the same ID number.

Table I. Distribution of the Duplicate ID Numbers

no. of carbon atoms	no. of alkane isomers	no. of duplicate ID numbers
15	4 347	1
16	10 359	1
17	24 894	3
18	60 523	8 (+1 triple)
19	148 284	23
20	366 319	88

Some of these are shown in Figure 1. In Table I we give the distribution of duplicates in each set of isomeric alkanes.

We searched the class of alkanes for counterexamples because the special structure of trees allows a very efficient method of computing the ID numbers (20 000 computer operations for a single alkane).

The obtained results may be summarized as follows: (1) The ID numbers are highly discriminating indices, but they are not unique. (2) There exist structures with very small differences in their ID numbers (see Figure 2). (3) For complicated structures (e.g., polyhexes) the ID numbers are not easily computed.[8]

In concluding this paper we point out that with this work we once again demonstrated the usefulness of developing generating algorithms that produce *all* members of a given family of (chemical) graphs and that thus make easier the check on many conjectures proposed in the field of (chemical) graph theory.

ACKNOWLEDGMENT

We thank Professor Milan Randić (Ames) for correspondence on ID numbers and for allowing us to examine his paper prior to the publication. We are also thankful to W. J. Wiswesser (Frederick, MD) and to the referees for their comments. This work was supported in part by the German-Yugoslav scientific cooperation program. The financial support by the Internationales Büro, Kernforschungsanlage Jülich, and by the Republic Council for Science of Croatia (SIZ II) is gratefully acknowledged.

REFERENCES AND NOTES

(1) Randić, M. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164.
(2) Trinajstić, N. "Chemical Graph Theory"; CRC Press: Boca Raton, FL, 1983.
(3) Harary, F. "Graph Theory"; Addison-Wesley: Reading, MA, 1972; third printing, Chapter 4.
(4) Knop, J. V.; Müller, W. R.; Jeričević, Ž.; Trinajstić, N. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 91.
(5) Trinajstić, N.; Jeričević, Ž.; Knop, J. V.; Müller, W. R.; Szymanski, K. *Pure Appl. Chem.* **1983**, *55*, 370.
(6) Knop, J. V.; Müller, W. R.; Szymanski, K.; Trinajstić, N. "Computer Generation of Some Classes of Molecules"; SKTH Press: Zagreb, Yugoslavia, 1985.
(7) Lang, S. "Algebra"; Addison-Wesley: Reading, MA, 1969; p 175.
(8) We carried out some computations of ID numbers for polyhexes. In this case we found that a prohibitive amount of computer time is needed. For example, in order to calculate the ID number of a polyhex with 10 rings, we needed 240 million computer operations (300 s of CPU time on a 0.8 MIPS computer). Thus, it was impossible to carry out the same analysis for the class of polyhexes.

End-User Searching: The Amoco Experience†

ROBERT E. BUNTROCK* and ALDONA K. VALICENTI
Amoco Research Center, Amoco Corporation, Naperville, Illinois 60566

Received December 18, 1984

The history of training scientists and engineers to do their own online searching of technical information is reviewed briefly. Searching services have been provided at Amoco for decades, computer searching for 15 years, and online services since 1973. In late 1981, it became apparent that several Amoco scientists and engineers wanted to learn to do at least some of their own online searching. We viewed this phenomenon as an opportunity, not a threat, and have provided training and assistance to our active, growing group of end-users since early 1982. We have divided the training into two parts: classroom background for local information resource awareness and individualized, hands-on training in online searching techniques. The program seems to be quite successful in retention of trained end-users, and most, if not all, participants seem to have significantly enhanced awareness of searching technical information.

"Can I do my own computer searching?" "Should I do my own computer searching?" Chemical information specialists have occasionally been asked these questions for the last decade, and the frequency is increasing. There seems to be no single reason for the increase in curiosity by the chemist on this subject, but greater familiarity with computers and computing is probably most important. Using titles similar to "The Library at Your Fingertips", authors in the recently erupted microcomputer trade press would have you believe that everyone will soon have their own microcomputer and will do all of their own searching. Although we do not believe this extreme scenario will happen, we do believe that end-user searching by at least some chemists and chemical engineers is here to stay. We will present a brief history of end-user searching and will then describe our experiences in training end-users at Amoco Corp. and the Amoco Research Center.

First, some definitions. "End-user searching" has come to mean literature searching by the eventual recipient of the information, namely, the customer, client, the expert (or the would-be expert), or decision maker. This "end-user", if an expert in a technical subject, is usually in research. Computer-based searching, specifically online, is implied.

Apparently, end-user searching appeared on the scene shortly after the beginning of readily available online searching in general, which, by our definition, is July 1973. We recall end-users in training classes for online systems or data bases as early as 1974-1975. By 1976, a paper had appeared titled "Nonmediated Use of Medline and Toxline by Pathologists and Pharmacists",[1] and additional work was cited. In 1977, Charles Meadow addressed the topic,[2] and in 1979 he published the pivotal paper of the field.[3] Meadow briefly reviewed the history of programming, and more specifically programmers, and then compared the two with online searching and search intermediaries on nine key issues. It is not our intent to summarize the paper extensively, but some quotes are very interesting. Programmers and intermediaries are described as having "...the keys to the kingdom..." and are often found

† Presented, in part, before the Division of Chemical Information, "Symposium on Training Chemists To Do Their Own Computer Searching", 187th National Meeting of the American Chemical Society, St. Louis, MO, April 11, 1984; American Chemical Society: Washington, DC, 1984; CINF 32; and "Symposium on Direct End-User Access to Chemical Information", 3rd Joint Meeting of the ACS Great Lakes Region and Central Region, Western Michigan University, Kalamazoo, MI, May 25, 1984; American Chemical Society: Washington, DC, 1984; paper 154.

0095-2338/85/1625-0415$01.50/0