6. Chvatal, V., Klarner, D.A., and Knuth, D.E. Selected combinatorial research problems. Tech. Rep. STAN-CS-72-292, Computer Sci. Dep., Stanford U., 1972.
7. Fletcher, J.G. A program to solve the pentomino problem by the recursive use of macros. *Comm. ACM 8*, 10 (Oct. 1965), 621–623.
8. Gardner, M. Mathematical games. *Scientific American*, Sept. 1966 and Jan. 1967.
9. Golomb, S.W., and Baumert, L.D. Backtrack programming. *J. ACM 12*, 4 (Oct. 1965), 516–524.
10. Hall, M., and Knuth, D.E. Combinatorial analysis and computers. *Amer. Math. Mo. 72*, 2 (Part II) (Feb. 1965), 21–28.
11. Knuth, D.E. *The Art of Computer Programming, Volume I: Fundamental Algorithms*, 2nd ed. Addison-Wesley, Reading, Mass., 1973.
12. Knuth, D.E. Estimating the efficiency of backtrack programs. *Math. Comp. 29* (1975), 121–136.
13. Kraitchik, M. *Mathematical Recreations*. W.W. Norton, New York, 1942; Revised ed., Dover, New York, 1953.
14. Lawler, E.L., and Wood, D.E. Branch-and-bound methods: a survery. *Oper. Res. 14* (1966), 699–719.
15. Lin, S. (personal communication).
16. Lucas, E., *Récréations Mathématiques*, 2nd ed. Gauthier-Villars, Paris, 1891.
17. Nievergelt, J. (personal communication).
18. Peterson, G. (personal communication).
19. Preparata, F., and Nievergelt, J. Difference-preserving codes *IEEE Trans. on Inf. Theory 20* (1974), 643–649.
20. Sainte-Laguë, M.A. *Les Reseaux (ou Graphes)*. Memorial des Sciences Mathematiques, Fasc. 18, Gauthier-Villars, Paris, 1926.
21. Slagle, J.R. *Artificial Intelligence—The Heuristic Programming Approach*. McGraw-Hill, New York, 1971.
22. Tarjan, R.E. Depth first search and linear graph algorithms. *SIAM J. Comput. 1* (1972), 146–160.
23. Tutte, W.T. The quest of the perfect square. *Amer. Math. Mo. 72*, 2 (Part II) (Feb. 1965), 29–35.
24. Walker, R.J. An enumerative technique for a class of combinatorial problems. *Combinatorial Analysis (Proceedings of Symposium in Applied Mathematics, Vol. X)*, Amer. Math. Soc., Providence, R.I., 1960.

Programming Techniques

G. Manacher, S. Graham Editors

A Note on the Set Basis Problem Related to the Compaction of Character Sets

Lawrence T. Kou and C.K. Wong
IBM Thomas J. Watson Research Center

This note discusses the reduction of the set basis problem to the clique cover problem.

Key Words and Phrases: compaction of character sets, set basis, set covering, computational complexity, polynomial completeness, clique cover

CR Categories: 4.9, 5.25, 5.39

In his paper on the minimization of spatially multiplexed character sets [2], Gimpel considered the following set basis problem. Given a collection of sets $S = \{S_1, S_2, \ldots, S_l\}$, a basis B is defined as a collection of sets, $B = \{B_1, B_2, \ldots, B_m\}$ such that for each S_i in S there exists a (possibly trivial) subset of B whose union equals S_i, the problem is to find a basis of least cardinality.

This important problem also arises in feature extraction and other areas in picture processing [1]. In most cases, the amount of computation required to solve this problem is prohibitively large.

In [2], the conversion of this problem to the set covering problem is discussed and can be described as follows. The columns of the covering problem correspond to element instances within the sets $\{S_i\}$. Thus the total number of columns is equal to the sum of the cardinalities of the $\{S_i\}$. For element e_j in set S_i, there

Authors' address: IBM Thomas J. Watson Research Center, P. O. Box 218, Yorktown Heights, NY 10598.

Table I.

	S_1		S_2		S_3		
	a	b	b	c	a	b	c
a	1	0	0	0	1	0	0
b	0	1	1	0	0	1	0
c	0	0	0	1	0	0	1
ab	1	1	0	0	1	1	0
ac	0	0	0	0	1	0	1
bc	0	0	1	1	0	1	1
abc	0	0	0	0	1	1	1

is a column which we may designate S_i/e_j. The rows of the table are sets which are candidates for membership in a basis $\{B_i\}$. A row representing set R_k covers a column S_i/e_j if and only if $e_j \in R_k \subseteq S_i$. Thus, if $S_1 = \{a, b\}$, $S_2 = \{b, c\}$, and $S_3 = \{\text{a, b, c}\}$, the covering problem can be represented by Table I.

As pointed out in [2], this conversion to a set covering problem has the advantage that many well-known methods to solve the set covering problem exist and can therefore be applied immediately. But it also has the disadvantage that the resulting matrix is too large in size. For example, if the number of distinct elements is n, the number of rows in the covering table is $2^n - 1$, corresponding to all the subsets of n objects except the void set. Note that this is only the worst possible case. In fact, there are usually fewer candidate sets because of the observations made by Gimpel in [2]. Therefore in the worst possible case the number of rows may grow exponentially with n. (The number of columns is $\alpha = |S_1| + |S_2| + \cdots + |S_l|$, where $|S_i|$ is the cardinality of S_i).

We show here that the set basis problem can be reduced to yet another well-known problem, namely, the clique cover problem. Notice that as far as time complexities are concerned, the set covering problem and the clique cover problem have been shown to be in the same class, i.e. both are polynomially complete [3]. However, our method reduces the set basis problem to the clique cover problem in polynomial time and the corresponding graph has $\alpha = |S_1| + |S_2| + \cdots + |S_l|$ nodes, hence at most $\binom{\alpha}{2}$ edges. Consequently, its size grows only polynomially with n.

First we need some definitions: given an (undirected) graph G, a *clique* is a complete subgraph of G; a *clique cover* of size k for G is a family of k cliques such that every node in G is in at least one of the cliques.

Next we show that given S one can construct a graph G such that any basis for S corresponds to a clique cover for G with the same cardinality and vice versa.

Given $S = \{S_1, S_2, \ldots, S_l\}$, where $S_i = \{i_1, i_2, \ldots, i_{k_i}\}$ for $i = 1, 2, \ldots, l$, consider the graph G_S constructed as follows. The set of nodes in G_S has a one-to-one correspondence to the elements of the set

$$X = \bigcup_{i=1}^{l} \{(S_i, i_j) \mid j = 1, 2, \ldots, k_i\}.$$

An edge is drawn between the two nodes corresponding to the elements (S_u, u_α) and (S_v, v_β) in X if and only if $S_u \cap S_v \supseteq \{u_\alpha, v_\beta\}$. The following proposition gives the relationship between the set basis problem for S and the clique cover problem for G_S.

PROPOSITION. *S has a basis of cardinality k if and only if G_S has a clique cover of size k.*

PROOF. Suppose S has a basis of cardinality k, say, $B = \{B_1, B_2, \ldots, B_k\}$. Consider the subgraph G_i of G_S with nodes corresponding to the set $X_i = \{(S_y, x) \mid S_y \supseteq B_i \text{ and } x \in B_i\}$ and with all the edges connecting nodes in X_i. By the construction of G_S, G_i forms a clique. Furthermore, since B forms a basis, for every element $(S_i, i_j) \in X$, there exists at least one basis element B_h such that $i_j \in B_h \subseteq S_i$. Therefore (S_i, i_j) is in G_h. Consequently, every element in X is in at least one of the cliques. Hence $G = \{G_1, G_2, \ldots, G_k\}$ forms a clique cover for G_S.

On the other hand, if G_S has a clique cover of size k, let $G = \{G_1, G_2, \ldots, G_k\}$ designate the clique cover. Define $B_j = \{x \mid (S_y, x) \text{ is a node in } G_j\}$, for $j = 1, 2, \ldots, k$. Notice that since G_j is a clique, for all (S_y, x) in G_j, $S_y \supseteq B_j$. Now, for each $S_i \in S$, let $T_i = \{s \mid G_s \text{ contains a node } (S_i, i_j), 1 \leq j \leq k_i\}$. Then

$$S_i = \{x \mid (S_i, x) \in X\} = \bigcup_{j=1}^{k} \mid \{x\, (S_i, x) \in G_j\}$$
$$= \bigcup_{j \in T_i} B_j.$$

Hence $B = \{B_1, B_2, \ldots, B_k\}$ forms a basis. □

Finally, it should be pointed out that Stockmeyer [4] recently proved the polynomial completeness of the set basis problem. Therefore it may be useful to study heuristics which yield near optimal results.

Received November 1974; revised March 1975

References

1. Block, H.D., Nilsson, N.J., and Duda, R.O. Determination and detection of features in patterns. In *Computer and Information Sciences*, J. T. Tou and R. H. Wilcox (Eds.), Spartan Books, Washington, D.C., 1964, pp. 75–110.
2. Gimpel, J.F. The minimization of spatially-multiplexed character sets. *Comm. ACM 17*, 6 (June 1974), 315–318.
3. Karp, R.M. Reducibility among combinatorial problems. In *Complexity of Computer Computations.* R.E. Miller and J.W. Thatcher (Eds.), Plenum Press, New York, 1972, pp. 85–103.
4. Stockmeyer, L. J., The minimal set basis problem is NP-complete. IBM Research Rep. RC 5431, May 1975.