

СРАВНЕНИЕ ПРОГРАММ ОБРАБОТКИ СТРУКТУРНОЙ  
ИНФОРМАЦИИ

А.А. Добрынин

Для обработки структурной информации, в частности графов, существует большое количество программ, число которых продолжает увеличиваться. Как правило, для решения конкретной прикладной задачи существуют наборы программ с похожими характеристиками. В данной работе экспериментально изучались свойства и возможности программ, предназначенных для решения следующих задач: распознавание изоморфизма графов, определение орбит группы автоморфизмов графа, нахождение группы автоморфизмов графа, получение канонического представления графов. Сравнительные исследования проводились для следующих программ: **ORBIT**, **GLICAN**, **SANGLOR**, созданные в ИМ СО АН СССР, Новосибирск [1,2]; программный комплекс **НОС** (**HIERARCHICALLY ORDERED EXTENDED CONNECTIVITY**), разработанный коллективом авторов из Болгарии и Румынии [3]; программная система **АУТОИЗОМ**, созданная в МЭИ, Москва [4]; программа **ГОТЛ**, разработанная в Институте органического синтеза ЛатвССР, Рига [5]; программные комплексы **АВТОГРАФ** и **ИЗОГРАФ**, созданные в МГУ, Москва [6]. Предполагается, что результаты исследования будут представлять интерес как для разработчиков, так и для пользователей — для первых в силу сравнения их программ с подобными, для вторых — в силу иллюстрации возможностей программ для их практического использования.

## 1. Описание программ

Ниже будет приведено описание исследуемых программ по следующей схеме: назначение, вид обрабатываемых графов (обыкновенный граф, мультиграф и т.п.), идеи алгоритма, теоретическая оценка

трудоемкости алгоритма (через  $p$  будет обозначаться порядок графа), длина представления исходной информации (на примере кодирования простого цикла  $C_{10}$  с последовательной нумерацией вершин и одинаково помеченными вершинами).

Программа ORBITM предназначена для определения орбит группы автоморфизмов графа. Обрабатываются помеченные мультиграфы. В основе работы алгоритма лежит применение процедуры из [7], заключающейся в разбиении вершин графа на классы эквивалентностей. Пусть есть некоторое текущее разбиение вершин графа. Для каждой вершины определяется вектор,  $i$ -я компонента которого равна числу вершин в  $i$ -м классе разбиения, смежных с рассматриваемой вершиной. В процессе разбиения учитывается кратность ребер. Вершины, имеющие одинаковые вектора, выделяются в новый класс. Процесс продолжается до стабилизации разбиений. На начальном этапе работы программы проводится разбиение вершин на классы, определяемые метками вершин графа. Трудоемкость алгоритма составляет  $O(p^3)$ . Входная информация представляется на языке описания структурной информации ОГРА-30 [8]. Длина представления графа - 13 символов. Язык реализации - ПЛ/I.

Программа ORBIT [1] предназначена для определения орбит группы автоморфизмов графа. Обрабатываются непомеченные мультиграфы. Алгоритм основан на методах относительного анализа графов [9]. Для каждой вершины графа подсчитывается набор частичных степеней, которые имеет данная вершина во всех относительных разбиениях. Вершины, имеющие одинаковые наборы частичных степеней, образуют классы эквивалентности. Далее применяется процедура из [7]. Трудоемкость алгоритма составляет  $O(p^3)$ . Входная информация представляется на языке ОГРА-30. Язык реализации - ПЛ/I.

Программа GLICAN [2] предназначена для распознавания изоморфизма графов. Обрабатываются помеченные мультиграфы. Сущность алгоритма заключается в итеративном применении процедуры из [7] и принудительном выделении произвольной вершины из неединичного класса в отдельный класс при стабилизации разбиения. В процессе разбиения классы вершин упорядочиваются. При несовпадении на каком-либо шаге соответствующих классов разбиения для двух графов определяется неизоморфность графов. Трудоемкость алгоритма составляет  $O(p^3)$ . Входная информация представляется на языке ОГРА-30. Язык программирования - ПЛ/I.

Программа SANGIOR [2] предназначена для построения канонического представления графа. Обрабатываются помеченные мультигра-

фы. В алгоритме производится итерационное разбиение вершин графа процедурой из [7] и принудительное выделение вершины в отдельный класс при стабилизации разбиения. Интересной особенностью программы является возможность получения канонического представления графа не в виде матрицы связей, а в виде линейной канонической записи графа на языке ОГРА-30 [2]. Трудоемкость алгоритма составляет  $O(p^3)$ . Входная информация представляется на языке ОГРА-30, язык реализации - ПЛ/I.

Программный комплекс НОС [3] предназначен для определения орбит группы автоморфизмов графа, канонической нумерации и получения линейного представления графа (топологический или химико-топологический код). Комплекс позволяет обрабатывать помеченные мультиграфы. В алгоритме используются идеи алгоритма Моргана [10], дополненные новыми положениями и принципами (сверхрасширенная связность, иерархичность разбиения и т.п.). Трудоемкость основного алгоритма составляет  $O(p^3)$ . Входная информация представляется в специальном виде. Длина представления составляет 220 символов. Язык реализации - ФОРТРАН.

Программная система AUTOIZOM [4]. Программы системы предназначены для определения орбит группы автоморфизмов графа, группы автоморфизмов, распознавания изоморфизма графов. Программы обрабатывают простые графы. Основные идеи алгоритмов - применение инвариантов, итеративное использование процедуры из [7], принудительное выделение произвольной вершины в единичный класс при стабилизации разбиения. Трудоемкость алгоритмов  $O(p^3)$ , определения группы автоморфизмов  $O(p^3 \cdot |AUT|)$ , где  $|AUT|$  - мощность группы автоморфизмов. Входная информация представляется в специальном виде. Длина представления составляет 120 символов. Язык реализации - ФОРТРАН.

Программа POTL предназначена для определения орбит группы автоморфизмов графа. Обрабатываются обыкновенные графы. В основе алгоритма лежит метод потенциалов [5]. Сущность метода заключается в рассмотрении графа как электрической схемы, в узлах которой устанавливаются под действием электрического тока некоторые потенциалы. Из физических соображений потенциалы должны характеризовать топологическую структуру схемы следующим образом: эквивалентные вершины имеют одинаковые потенциалы. Потенциалы определяются из системы линейных уравнений. Трудоемкость алгоритма  $O(p^3)$ . Входной информацией является матрица смежности графа. Язык реализации - ФОРТРАН.

Программные комплексы АВТОГРАФ и ИЗОГРАФ [6] предназначены для определения группы автоморфизмов графа, канонической нумерации и распознавания изоморфизма. Обрабатываются помеченные мультиграфы. Идея алгоритмов заключается в построении так называемого дерева подстановок. Для каждой строящейся подстановки проверяются некоторые соотношения, называемые межцикловыми и внутрицикловыми критериями, которые являются достаточными для определения автоморфности подстановки. В алгоритме предусмотрено сокращение перебора путем удаления некоторых поддеревьев дерева подстановок. Оценка трудоемкости алгоритма в [6] не приводится. Входная информация представляется в специальном виде. Длина представления составляет 150 символов. Язык реализации - ФОРТРАН.

## 2. Результаты сравнения

Экспериментальное исследование программ включало выяснение в процессе практического использования ряда их характеристик. Основное внимание было обращено на получение временных характеристик и определение графов, на которых программы дают неверные результаты (контрпримеры к алгоритмам). Для изучения поведения программ были выбраны следующие классы графов: I - кубические графы с тождественной группой автоморфизмов (порядок графов от 12 до 40 вершин), II - полициклические графы (конденсированные шестицикленные кольца, число колец от I до 10, порядок графов от 6 до 32 вершин), III - транзитивные графы с большим порядком группы автоморфизмов (порядок графов от 4 до 16 вершин), IV - специально помеченные графы, V - сильно регулярные графы. Кубические графы были выбраны жесткими, так как подавляющее число графов химических соединений имеют тождественную группу автоморфизмов или обладают малой степенью симметрии. Полициклические графы также хорошо отражают химическую реальность. Класс III взят в силу существенных затрат ресурсов для его обработки. Классы IV и V были выбраны для определения мощности алгоритмов, реализуемых данными программами. На рис. 1-6 приводятся временные характеристики программ на классах графов I-III, где T - время обработки графа (сек), P - порядок графа. Для наглядности графики представлены непрерывной линией. Программы на рисунках перечислены в порядке возрастания временных затрат. Критерием оценки структурной сложности графов, обрабатываемых программами, являлись классы IV и V. На рис. 7 приведен планар-

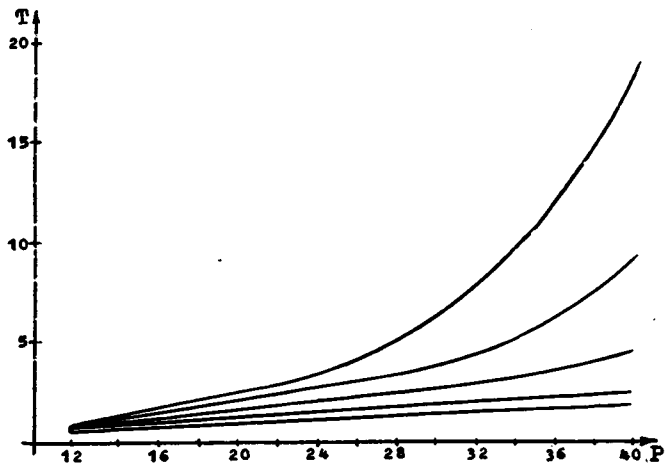


Рис. 1. Определение орбит для класса I (НОС, РОТЛ, АУТОИЗОМ, ОРНИМ, ОРВИТ).

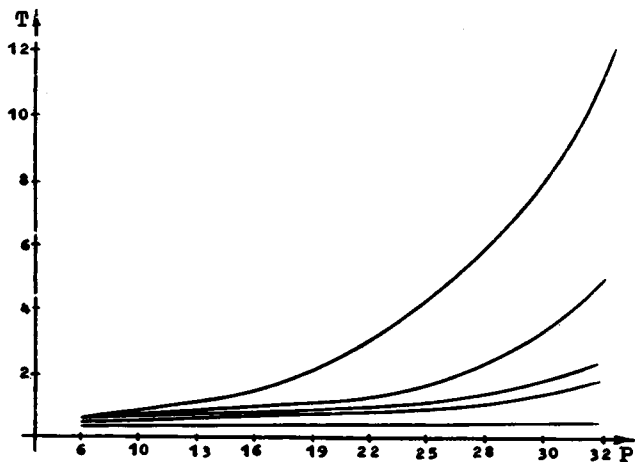


Рис. 2. Определение орбит для класса II (НОС, РОТЛ, АУТОИЗОМ, ОРНИМ, ОРВИТ).

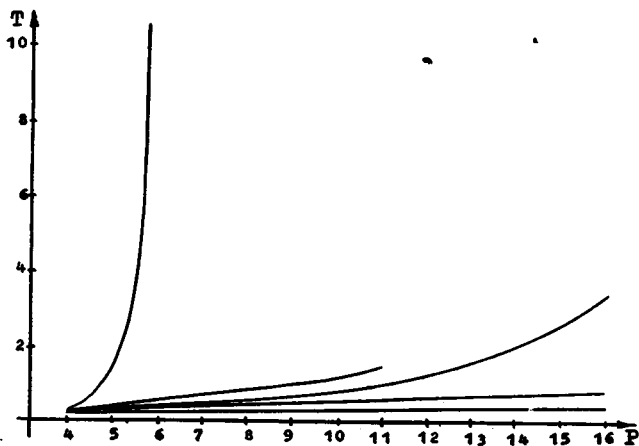


Рис. 3. Определение орбит для класса III, полные графы (POTL, ORNIM, OBBIT, HOC, AUTOIZOM).

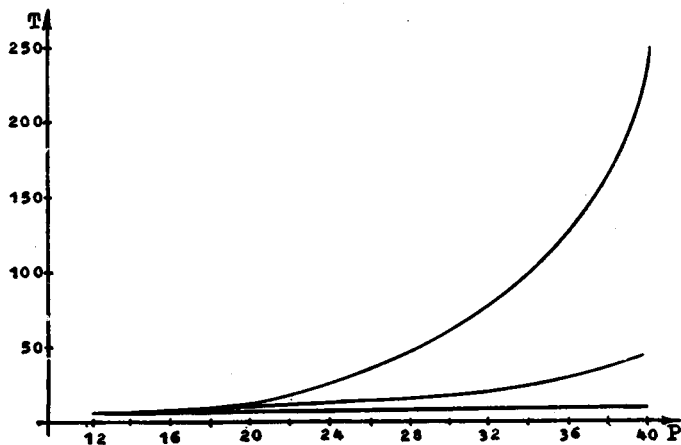


Рис. 4. Распознавание изоморфизма класса I (AUTOIZOM, GLICAM, ИЗОГРАФ).

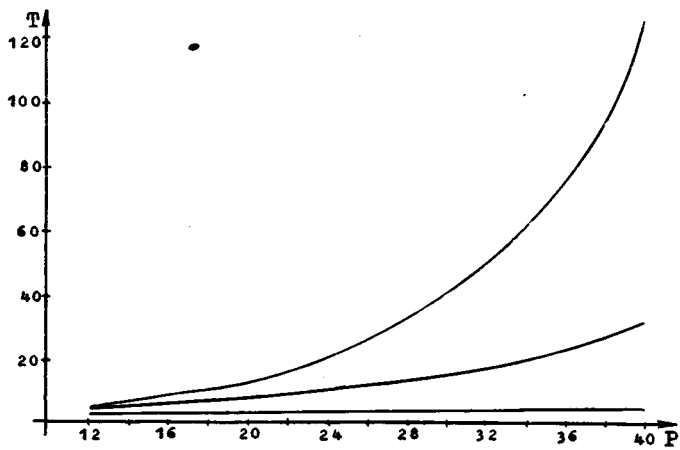


Рис. 5. Канонизация для класса I  
(НОС, GANGLOR, ИЗОГРАФ).

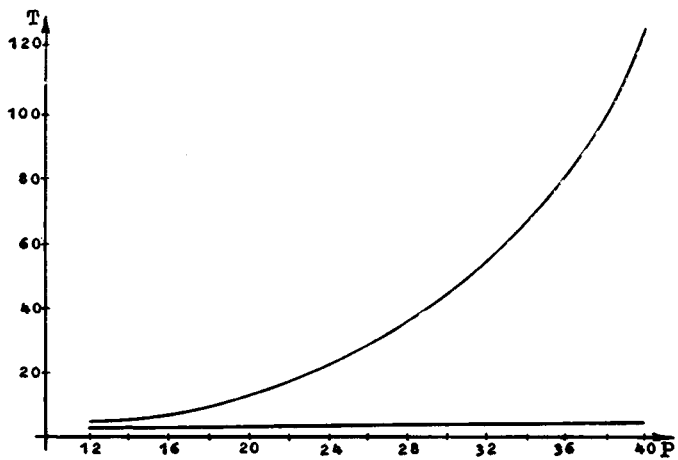


Рис. 6. Определение автоморфизмов для класса I  
(АУТОИЗОМ, АВТОГРАФ).

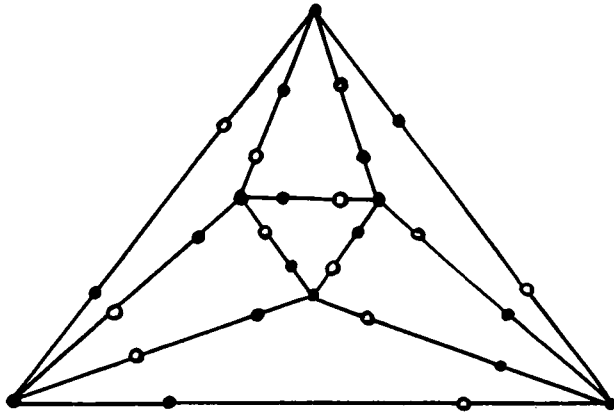


Рис. 7. Граф из класса IV.

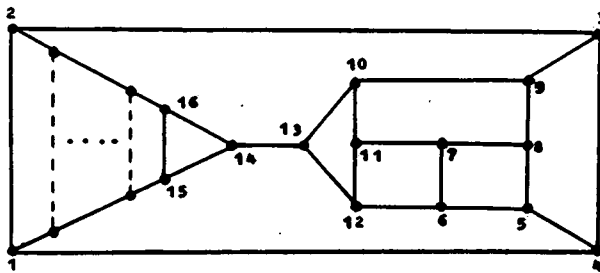


Рис. 8. Граф для метода потенциалов.



ный граф – представитель класса IV, вершины которого разбиваются на 14 орбит мощности 3. Граф наглядно иллюстрирует, например, неполноту такого инварианта, как эквивалентность путей и цикловых характеристик вершин. Графы класса IV являются контрпримерами к алгоритмам программ ORNIM, NOS, POTL, GLICAN, SLANGOR. Сильно-регулярные графы являются контрпримером к программе ORBIT.

Для метода потенциалов была отмечена одна особенность (в качестве компонент вектора свободных членов в системе линейных уравнений использовались величины, обратные диагональным элементам обратной матрицы системы). Суть ее заключается в том, что способность метода потенциалов на практике различать неэквивалентные вершины зависит не только от структурной сложности графа, но и просто от порядка графа. Для иллюстрации рассмотрим граф на рис.8 порядка 16, для которого программа POTL дает правильный результат (граф имеет тождественную группу автоморфизмов). Построим семейство графов, добавляя на ребра (1,15), (2,16) вершины и соединяя их ребрами, показанными на

Т а б л и ц а

Порядок графа	Потенциалы вершин I5 и I6
20	2,943883561412973 2,943884461532764
24	2,943562629976453 2,943562669362765
32	2,943508575239400 2,943508575314651
40	2,943507423146722 2,943507423146864
42	2,943507407659830 2,943507407659860
46	2,943507399484837 2,943507399484838
48	2,943507398621779 2,943507398621779

рис.8 пунктирной линией. При увеличении порядка графа (т.е. при добавлении новых ребер описанным выше способом) потенциалы вершин I5 и I6 становятся все менее различимыми, что хорошо заметно из таблицы. Их различие сдвигается к границе точности вычислений. При порядке графа 48 (при счете с двойной точностью) потенциалы вершин I5 и I6 совпали (из-за совпадения диагональных элементов обратной матрицы). Вычисления проводились с использованием программы, разработанной в [5] и адаптированной для ЕС ЭЕМ.

Все исследования программы проводились на ЭВМ ЕС I050. Ряд программ ввиду громоздкого представления исходной информации был модернизирован для возможности применения языка ОГРА-30 для представления графов.

## Л и т е р а т у р а

1. СКОРОБОГАТОВ В.А., ХВОРОСТОВ П.В. Анализ симметрий графов. - Тез. докл. П Всесоюз. совещ. "Методы и программы решения оптимизационных задач на графах и сетях". Ч.П. Новосибирск, 1982, с. 43-45.
2. ЗАГОРУЙКО Н.Г., СКОРОБОГАТОВ В.А., ХВОРОСТОВ П.В. Вопросы анализа и распознавания молекулярных графов на основе общих фрагментов. - Настоящий сборник, с. 26-50.
3. BALABAN A.T., BONCHEV D., MEKENYAN O. Ordering and numbering of atoms, and unique topological representation of chemical structures by algorithms. - Тез. докл. VI Всесоюз. конф. "Использование вычислительных машин в спектроскопии молекул и химических исследованиях", Новосибирск, 1983, с. 87-89.
4. БАШМАКОВ И.А., КОХОВ В.А. Пакет прикладных программ "Изоморфизм-Аutomорфизм". - Тез. докл. П Всесоюз. совещ. "Методы и программы решения оптимизационных задач на графах и сетях". Ч. I. Новосибирск, 1982, с. 21-24.
5. GOLENDER V.E., DRBOGLAV V.V., ROSENBLIT A.B. Graph Potentials Method and its Application for Chemical Information Processing. - J. Chem. Inf. Comput. Sci., 1981, v. 21, p. 196-204.
6. Алгоритм поиска всех подстановок группы автоморфизмов графа/ Жидков И.П., Зефирова М.С., Попов А.И. и др. - В кн.: Математические вопросы структурного анализа и алгоритмы машинных экспериментов в органической химии. 1979, с. 59-73.
7. CORNEIL D.G., GOTLIEB S.C. An efficient algorithm for graph isomorphism. - Journal ACM, 1970, v. 17, N 1, p. 51-64.
8. КОЧЕТОВА А.А., СКОРОБОГАТОВ В.А., ХВОРОСТОВ П.В. Язык описания структурной информации ОГРА-30. - В кн.: Машинные методы обнаружения закономерностей, анализа структур и проектирования (Вычислительные системы, вып. 92). Новосибирск, 1982, с. 70-79.
9. СКОРОБОГАТОВ В.А. Относительные разбиения и слои графов. - В кн.: Вопросы обработки информации при проектировании систем (Вычислительные системы, вып. 69). Новосибирск, 1977, с. 3-10.
10. MORGAN H.L. The generation of a unique machine description for chemical structures - a technique developed at Chemical Abstracts Service. - J. Chem. Doc., 1965, N 5, p. 107-113.

Поступила в ред.-изд. отд.

23 мая 1984 года