

УДК 681.142+002.6+576.8

Ր.Լ. ՄԱՐԿՕՅԱ

## НЕКОТОРЫЕ ПОДХОДЫ ПРИМЕНЕНИЯ МЕТОДОВ ИНФОРМАТИКИ В МИКРОБИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Исследована проблема применения методов и средств информационной технологии и вычислительной техники в микробиологии, в том числе в коллекциях культур микроорганизмов. Рассматривается вопрос о использовании специализированного прикладного программного комплекса для решения ряда фундаментальных и практических задач.

В последние годы значительное увеличение объемов биологической информации и развитие новых вычислительных методов привели к появлению нового научного направления – биоинформатике, что можно определить как научное направление, сформировавшееся на стыке вычислительной науки, биологии и математики. Применение методов информационной технологии привело к достижению значительных результатов во многих областях биологической науки [1].

В различных странах мира, в том числе и в Армении, в настоящее время созданы коллекции культур микроорганизмов (ККМ), которые служат в качестве депозитарий для микроорганизмов. Они одновременно являются источниками разносторонней информации о свойствах микроорганизмов, играющих важное значение для фундаментальной и прикладной микробиологии и биотехнологии [2]. Нарастающими темпами внедряются в эту область методы компьютерной обработки данных, их хранения, поиска и анализа. Применение информационной технологии предоставляет возможность для новых, качественно иных подходов и решений задач в области микробиологии и биотехнологии.

Методы вычислительной техники, распознавания образов, математической статистики, математического моделирования используются при решении вопросов систематики и классификации, генетики, экологии микроорганизмов, моделирования различных процессов с использованием автоматизированных систем их управления. Благодаря этому становится возможным создание и применение новых таксономических методов с учетом эволюции и филогении микроорганизмов, комплексного изучения различных свойств, признаков и процессов [3].

Исходя из вышеизложенного, мы задались целью создать некий компьютерный комплекс, позволяющий решать следующие фундаментальные и прикладные задачи в микробиологии:

- анализ и систематизацию данных, идентификацию и классификацию микроорганизмов;
- иерархическую, приоритетную компиляцию признаков;
- процент подобия, разброс данных.

Указанный комплекс будет функционировать на основе базы данных микроорганизмов. При функционировании системы происходит поиск вектора свойств, идентифицирующего данный микроорганизм. Эта идея находит свое развитие при присваивании различных весов свойствам микроорганизмов и в корне отличается от ныне принятого "адансоновского" метода, что основан на принципе равноправия всех признаков, используемых при кластеризации. Научный спор в отношении равноправия признаков при классификации биологических объектов имеет достаточно

долгую историю [4,5]. Фактически экспериментальный материал показывает, что во многих случаях целесообразно применять взвешивание (придание признакам неодинаковой "цены") в таксономических исследованиях микроорганизмов.

С учетом современных быстрых темпов развития вычислительной техники и программного обеспечения, указанная система открыта для дальнейшего усовершенствования и постоянного расширения.

В настоящее время почти во многих существующих базах данных коллекций культур микроорганизмов в качестве стандарта принята система РКС кодирования микробиологической информации для компьютеров [6,7]. В формате РКС все данные категорированы и маркированы для ввода их в компьютер системой шестизначных чисел, а в качестве основного метода для кодирования принята бинарная логика. РКС-код является незамкнутой (open-ended) системой, имеет возможность расширения и позволяет добавлять новые данные в систему кодирования. В нашем комплексе для кодирования информации с целью обеспечения совместимости также используется РКС-код. Однако надо отметить, что при использовании различных методов для "взвешивания" признаков исследуемых микроорганизмов желательно оперировать признаками, имеющими разные модальности.

Предложенный нами комплекс существенно отличается от ныне действующих аналогичных систем [8-11] как с точки зрения структуры, хранения и информационной взаимозависимости данных, так и с точки зрения функциональной деятельности.

В течение нескольких лет произошел также значительный прогресс в применении компьютерной графики в биологии [12]. В микробиологии эта область в первую очередь связана с микроскопией, все чаще использующей вычислительные системы. Необходимо подчеркнуть важность применения технологии обработки изображений в указанной сфере [13-15], поэтому нами принципиально предусмотрено хранение графической информации, в первую очередь, различных микроскопических снимков (световых и электронных), графических характеристик разнообразных процессов и др.

Кроме фенотипической информации, которая кодируется при помощи РКС – кода, будут храниться филогенетические деревья, номенклатурная база данных. В Институте микробиологии НАН РА долгие годы изучались проблемы биоразрушения различных синтетических полимеров специального назначения. Накоплен огромный информационный материал, представляющий собой большой научный и практический интерес. В дальнейшем система будет также включать в себя информацию о полимерных материалах, данные о взаимодействии в системе микроорганизм – полимер [16]. В существующей литературе не упоминаются какие-либо информационные, вычислительные системы, которые сохранили бы такого рода информацию.

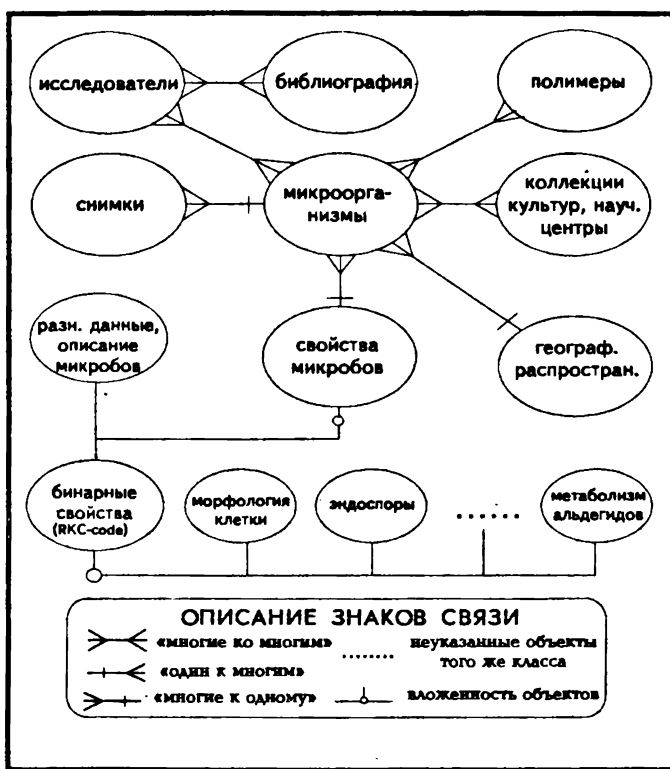
Целесообразно также интегрировать в предлагаемую систему созданные в настоящее время различные программные средства для идентификации, классификации разнообразных видов микроорганизмов и решения других задач.

В процессе развития компьютерной и информационной технологии очевидна тенденция построения разнообразных локальных, региональных и глобальных сетей. В Армении действуют несколько научно-образовательных и медицинских учреждений в области микробиологии и биотехнологии, коллекция культур микроорганизмов, санитарно-эпидемиологическая служба и другие организации, для которых посредством коммуникационных линий будет доступен данный микробиологический комплекс. Таким образом, предлагаемая система может стать реальной основой для создания республиканской микробиологической компьютерной сети – единственной в данном регионе, что существенно важно и необходимо для изучения экологии, биоразнообразия и других региональных и глобальных проблем. Кроме того, указанная сеть может оказать существенную помощь при решении медико-биологических и санитарно-эпидемиологических вопросов, часто требующих оперативную обработку значительных объемов информации, в частности в целях обеспечения эпидемиологической защиты.

Во многих странах проводятся исследования в области применения информа-

ционной технологии в микробиологии. Действуют специальные как национальные, так и международные организации для координации исследовательских программ и согласованной деятельности – Всемирная федерация коллекций культур (World Federation for Culture Collections, WFCC), Всемирный центр данных для микроорганизмов (World Data Centre for Microorganisms, WDCM) и другие. Созданы также микробиологические сети, напр., Сеть данных штаммов микроорганизмов (Microbial Strain Data Network, MSDN), Европейская сеть микробной информации (Microbial Information Network Europe, MINE). Создание микробиологической информационной сети в Армении предоставит возможность для естественной интеграции ее в глобальную сеть [17].

При проектировании системы обработки данных нас в первую очередь интересует их организация. С целью создания микробиологического компьютерного комплекса нами исследована взаимосвязь различных элементов и разработана информационная модель (см. рис. ). Определены и ниже представлены ее основные элементы



Информационная модель

и их взаимосвязь в информационном поле базы данных коллекции культур микроорганизмов. Необходимо заметить, что представленная информационная модель является упрощенной схемой и преследует цель определения лишь характерных взаимосвязей. Естественно, центральным является элемент "микроорганизмы" (регистрационный номер в БД и ККМ, наименование и т.д.), который может находиться в отношении "многие ко многим" с элементами "исследователи", "коллекции культур" и "полимеры". Элемент "снимки" (светового и электронного микроскопов) с элементом "микроорганизмы" связан отношением "многие к одному". Элемент "микроорганизмы" находится в отношении "многие к одному" с элементами "свойства микроорганизмов" и "географическое распространение".

Наиболее объемным и существенным является элемент "свойства микроорганизмов", который в данном случае состоит из двух частей: "бинарных свойств" и "разнообразных данных, описания микроорганизмов". Бинарные свойства основаны

на принципах кодирования микробиологической информации (РКС-код). Нашей главной целью являлось определение отношений между элементами, поэтому полностью не раскрыт элемент "бинарные свойства".

Для представления микробиологической информации в БД самой удобной является реляционная модель. Это в первую очередь определяется и сущностью метода кодирования микробиологической информации – РКС-кода.

Институт микробиологии НАН РА, ЕГУ

Поступила 20.05.1998

#### ЛИТЕРАТУРА

1. Trends in Biotechnology. 1996, v. 14, No.8 (151), p. 261-321.
2. Каталог культур микроорганизмов. Ер., Изд-во Гитутюн, НАН РА, 1996, с. 9-17.
3. Bryant T.N., Wimpenny J.W.T. Computers in microbiology. A practical approach. Oxford University Press, 1989.
4. Смирнов Е.С. Таксономический анализ. М.: Изд-во МГУ, 1969, с.9-17.
5. Песенко Ю.А. Принципы и методы количественного анализа в фаунистических исследованиях. М.: Изд-во Наука, 1982, с. 122-125.
6. Афрякян Э.Г., Котоян В.К., Хачатурян А.А. Дескрипторы для создания автоматизированного банка данных культур микроорганизмов. Ер., 1987.
7. Rogosa M., Krichevsky M.I., Colwell R.R. Coding Microbiological Data for Computers. New York, 1986.
8. Janke K.-D. Microbial strain data management and bacterial identification using IBM-compatible microcomputers. – Proceedings of the UNESCO-ROSTE European Workshop and Postgraduate Training Course "Modern Biotechnology". Prague, 1991, p. 248-269.
9. David A. Portyral. MICRO-IS: A Microbiological Database Management and Analysis System. Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision. Baltimore, Johns Hopkins University Press, 1993, p. 313-327
10. Ozerskaja S.M., Vasilenko A.N. Biomatrix 3.0 and Cluster Analysis program. Bynary, 1992, v. 4, pp.64-65.
11. Ozerskaja S.M., Vasilenko A.N. Biomatrix 3.0. Bynary, 1994, v. 6, p. 141-144.
12. Ramsom R., Matela R. J.. Computer Graphics in Biology. New York, 1986.
13. Маркосян Р.Л. Применение компьютерной обработки микроскопических изображений в базах данных коллекций культур. – Материалы конференции: Электронная микроскопия-97, Ер., 1997.
14. James F. Rohlf. Computer Vision Needs in Systematic Biology. Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision. Baltimore: Johns Hopkins University Press, 1993, p. 365-373.
15. Patrick A. D. Powell. Direction for Computing Research in Systematic Biology Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision. Baltimore: Johns Hopkins University Press, 1993, p. 449-458.
16. Mardjianian G., Simonyan M., Afrikian E. Approaches for creation of computer simulation and prediction system in "Microbial biodegradant-Synthetic Polymer" environment. Proceedings of the conf. Computer Science and Information Technologies, Yerevan, 1997.
17. The Integration of Microbial Databases. Report of a Workshop held at Michigan State University, 1995.

#### Ռ. Լ. ՄԱՐԿՈՍՅԱՆ

ՄԱՆՐԵԲԱԲԱՆԱԿԱՆ ՀԵՏԱԶՈՏՈՒԹՅՈՒՆՆԵՐՈՒՄ ԻՆՖՈՐՄԱՏԻԿԱՅԻ  
ՄԵԹՈԴՆԵՐԻ ԿԻՐԱՌՄԱՆ ՈՐՈՇ ՄՈՏԵՅՈՒՄՆԵՐ

#### Ա մ փ ո փ ու մ

Հետազոտված են տեղեկատվական տեխնոլոգիայի և հաշվողական տեխնիկայի մեթոդների և միջոցների կիրառման որոշ հիմնահարցերն ու հիմնախնդիրները մանրէաբանության մեջ, այդ թվում մանրէների հավաքածուներում: Առաջարկվում է օգտագործել հատուկ կիրառական ծրագրային համալիր հիմնարար և գործնական խնդիրների լուծման համար: