

ФОРМИРОВАНИЕ БАЗЫ ДАННЫХ ПО СОРТАМ ВИНОГРАДА РАЗЛИЧНОГО ЭКОЛОГО-ГЕОГРАФИЧЕСКОГО ПРОИСХОЖДЕНИЯ ДЛЯ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ АМПЕЛОЦЕНОЗОВ¹

Попова Д.В.

ООО «Малое инновационное предприятие «Ампелоинформпродукт»

Ильина И.А., *д-р техн. наук*

Петров В.С., *д-р с.-х. наук*, **Соколова В.В.**, *канд. с.-х. наук*

*Федеральное государственное бюджетное научное учреждение «Северо-Кавказский федеральный научный центр садоводства, виноградарства, виноделия»
(Краснодар)*

Реферат. В статье представлены результаты исследований по разработке базы данных хозяйственно-ценных признаков сортов винограда различного эколого-географического происхождения, включающей описание более чем по 80 признакам, как составной части биоинформационной технологии управления продуктивностью ампелоценозов и многоплановой, полифункциональной, гибкой информационной системы для хранения, извлечения информации о сортах, об экологических ресурсах, о сортовых технологиях возделывания винограда в конкретных почвенно-климатических условиях с учетом их назначения, необходимых сельхозтоваропроизводителям для принятия управленческих решений по созданию и эксплуатации ампелоценозов с целью обеспечения конкурентоспособности производимой продукции.

Ключевые слова: сорта винограда, хозяйственно-ценные признаки, компьютерное моделирование, база данных, словарь данных, структура базы данных

Summary. The article presents the results of studies on the development of a database of economically valuable traits of grape varieties of various ecological and geographical origin, including a description of more than 80 characters, as an integral part of bioinformation technology for managing the productivity of ampeloceneses and a multifaceted, multifunctional, flexible information system for storing, extracting information about varieties, about ecological resources, about varietal technologies for the cultivation of grapes in specific soil and climatic conditions taking into account their purpose, necessary for agricultural producers to make managerial decisions on the creation and operation of ampeloceneses in order to ensure the competitiveness of manufactured products.

Keywords: grape varieties, economically valuable traits, computer modeling, database, data dictionary, database structure

Введение. Современная концепция стабильного конкурентоспособного производства винограда основана на формировании устойчивых саморегулирующихся высокоадаптивных агроценозов, рациональном природопользовании, улучшении среды произрастания винограда [1].

¹ Исследование выполнено при финансовой поддержке РФФИ и администрации Краснодарского края в рамках научного проекта № 19-44-230021 p_a

В последние десятилетия на Юге России в следствии нестабильных погодных условий средний уровень реализации потенциала хозяйственной продуктивности используемых сортов винограда составляет 60 %, что в сочетании с технологиями, отличающимися от западноевропейских большей сложностью, приводит к снижению конкурентоспособности отечественного виноградарства [2].

Для решения этой задачи актуальным является разработка методов управления ростовыми и продукционными процессами насаждений, средообразующим потенциалом ампелоценозов на основе использования современных биоинформационных (цифровых) технологий и систематизации научных знаний о хозяйственно-ценных признаках винограда, влиянии природных и антропогенных факторов на реализацию потенциала хозяйственной продуктивности.

Одним из основных этапов такой работы является подбор высокоадаптивных сортов, формирование баз данных, содержащих информацию по биологическим и хозяйственно-ценным признакам сортов винограда.

Сбор и хранение информации о сортименте проводится во многих странах, выращивающих виноград: *SwissVitisMicrosatelliteDatabase* (SVMD, швейцарская база данных микросателлитных маркеров винограда); *GreekVitisdatabase* (греческая база данных *Vitis* L.); *Italian Vitis Database (VitisDB)* (публично открытая итальянская база данных *Vitis* L.); *Portuguese Vitis Database* (португальская база данных *Vitis* L.); *La Vigne et le vin* (французская база данных, отличается от указанных выше тем, что помимо информации о сортименте содержит описание областей возделывания Франции); *Vitis International Variety Catalogue* (международный каталог винограда, содержит информацию о более чем 1500 сортах со всего мира).

Все вышеперечисленные разработки представляют собой, по сути, электронные хранилища статистических данных, при этом нет моделей, которые на основе включенных данных способны сгенерировать рекомендации по уходу за существующими виноградниками и закладке новых. Кроме того, описание сортов производится по 7-20 параметрам.

Программы, используемые в других странах, производящих виноградо-винодельческую продукцию не могут быть использованы в РФ, т.к. имеют сжатую и ограниченную информацию о российском сортименте, не учитывают ресурсный потенциал зон возделывания винограда нашей страны, происходящие изменения климата и отдельные элементы технологий.

В Российской Федерации накоплен уникальный коллекционный фонд автохтонных, интродуцированных и стародавних сортов, а также современной селекции различного видового происхождения. Однако отсутствует точная и системная информация по широкому спектру адаптивности, экологической устойчивости, биологическому и продукционному потенциалам сортов и форм в разных природных, почвенно-климатических условиях географически удаленных друг от друга регионов. Кроме того, нет системных моделей расчета базовых и оперативных технологий. Расчет отдельных элементов технологий (орошение, питание, нагрузка куста урожаем, длина обрезки лоз, защита от болезней и вредителей и др.) зачастую проводится разными лицами (организациями) и их рекомендации могут не только быть плохо совместимыми, но и могут взаимоисключать друг друга.

Таким образом, в настоящее время, для условий нашей страны необходима разработка биоинформационных технологий и компьютерной программы, включающей обобщенные статистические данные об ампелографических коллекциях, ресурсном потенциале областей возделывания и отдельных хозяйств, которые с помощью созданных моделей расчета базовых и оперативных технологий на основе мониторинга среды, диагностики и накопленной информации позволят генерировать рекомендации по управлению продуктивным потенциалом ампелоценозов. Для получения максимальной эффективности от использования такой компьютерной модели необходимо заполнение баз данных информа-

цией о сортах. Причем, важно получить и внести в базу не только статистическую информацию, но и качественные показатели сортов в различных климатических зонах с целью подбора для каждой зоны оптимального сортимента, реализующего именно в этих почвенных и климатических условиях весь свой продукционный потенциал. На основе этих данных можно не только делать выводы об успешности отдельных ампелоценозов, но и давать рекомендации для повышения их эффективности. Кроме того, на основе собранных баз данных можно построить самообучающуюся систему, которая позволит подбирать оптимальный сортимент и агротехнологические приемы не только для конкретных почвенно-климатических условий, но и учитывать изменение климата в динамике.

Объекты и методы исследований. Объекты исследований – сорта винограда различного эколого-географического происхождения. Источником информации о сортах винограда являлись: Всероссийская ампелографическая коллекция (держатель – Анапская зональная опытная станция виноградарства и виноделия – филиал ФГБНУ Северо-Кавказского федерального научного центра садоводства, виноградарства, виноделия, г. Анапа), насчитывающая 4879 образцов винограда [11]; коллекция НИВиВ «Магарач» в Крыму, насчитывающая 3270 сортов и форм винограда; коллекция ВНИИВиВ им. Я.И. Потапенко (Ростовская область) – 1120 сортов и форм винограда; Дагестанская коллекция в г. Дербенте, насчитывающая 368 сортов. Основной упор при сборе информации сделан на показатели, характеризующие устойчивость и продуктивность виноградного растения (засухоустойчивость, морозостойкость, зимостойкость и др.) [3].

В качестве методологической базы исследований использовались современные методы моделирования и информационных технологий. При проведении исследований использовался комплексный подход к решению поставленных задач, характеризующийся использованием возможностей и методов информационных технологий, математического анализа. Для составления баз данных использовались также эмпирические данные, ранее полученные в Северо-Кавказском федеральном научном центре садоводства, виноградарства, виноделия.

Процесс информационного изучения сортов включает 3 основные фазы: аналитическую, причем для повышения точности оценки используется двухступенчатый подход – результаты полевых наблюдений за реакцией растений на изменения среды и лабораторная оценка объектов в моделируемых условиях эксперимента; синтетическую (статистический и математический анализ полученных данных) и обобщающую (результатирующую оценку образца, получение информации о взаимодействии образца со средой на основании результатов моделируемых стрессовых воздействий, выработку рекомендаций по использованию образца в той или иной климатической зоне) [4-5].

Обсуждение результатов. Целью проведенных исследований являлось формирование базы данных хозяйственно-ценных признаков сортов винограда на основе агробиологических и физиолого-биохимических результатов исследований генотипов.

База данных – программа, позволяющая получать и хранить большие объемы связанной информации. Она состоит из таблиц, содержащих информацию [6]. На этапе ее создания необходимо определить какие таблицы нужно создать и какие связи будут существовать между информацией в таблицах, т.е. разработать проект базы данных. Хороший проект базы данных обеспечит целостность данных и простоту их обслуживания.

Основные задачи проектирования баз данных:

- обеспечение хранения в БД всей необходимой информации;
- обеспечение возможности получения данных по всем необходимым запросам;
- сокращение избыточности и дублирования данных;
- обеспечение целостности базы данных [7].

Разработка эффективной базы данных состоит из нескольких этапов:

– процесс разработки БД начинается с анализа требований, то есть необходимо определить: какие элементы данных должны храниться в базе, кто и как будет к ним обращаться;

– создается логическая структура БД, то есть определяется, как данные будут сгруппированы логически;

– логическая структура БД преобразуется в физическую с учетом аспектов производительности. Элементы данных на этом этапе получают атрибуты и определяются как столбцы в таблицах выбранной для реализации БД СУБД [8].

Основная сложность проектирования базы данных – необходимо придумать представление для реальных объектов и их связей в терминах таблиц, полей, атрибутов, записей, то есть в терминах абстракций реляционной модели данных.

База данных для прототипа программы управления продуктивностью ампелоценозов – представляет собой группу связанных таблиц с описанием сортимента, областей возделывания и опытных хозяйств и вспомогательных таблиц для хранения информации, используемой для расчета базовых и оперативных технологий, оценки эффективности действующих насаждений, прогнозов очагов вспышек болезней и вредителей, выявления зон для возделывания виноградников.

Создание базы данных ценных признаков сортов проходило в три этапа.

Этап 1. Анализ требований

Общая методика определения требований к БД заключается в составлении словаря данных. Словарь данных перечисляет и определяет отдельные элементы данных, которые должны храниться в базе.

Важная часть анализа требований – предупредить потребности пользователей, то есть словарь данных должен содержать информацию, которая описывает сорта как можно более полно и подробно [9].

На этапе создания словаря можно уже определить тип данных для каждого элемента (в последствии – столбца), это сэкономит время проектирования.

Помимо хранения данных, необходимо будет извлекать их из базы, использовать в алгоритмах расчета технологий, поэтому текстовые элементы таблицы, желательно заменить на числовые, так как программному коду удобнее обрабатывать числа, а не текст. Поэтому, при составлении словаря продуманы, какие текстовые характеристики можно заменить на числовые.

В качестве параметров, наиболее полно описывающих хозяйственно-ценные признаки сортов, выбрана 81 характеристика, это выгодно отличает разработку от мировых аналогов (описание сортов производится 7-20 параметрами), так как дает возможность пользователю получить более детальную характеристику сорта.

Разбивка одного параметра на несколько унифицированных (например, описание ягоды разбито на 5 характеристик: форма, размер, вес, цвет, наличие семян) дает возможность быстрого поиска и подбора сортов (вместо поиска в тексте значения слова «бессемянной», можно просто отфильтровывать сорта по ячейке «наличие семян»).

Для определения всех необходимых характеристик сорта были проанализированы по требованиям, предъявляемым различными пользователями к разрабатываемой системе, в следствии чего, все данные, описывающие сорт, разделены по функциональным группам.

Основной параметр для описания сорта винограда – его наименование (*Сорт*) – это уникальное наименование. Не существует двух сортов с одним и тем же названием. Тип данных – текстовый, длина строки – 50 символов.

Общая характеристика сорта – информация о сорте, которая интересна всем видам пользователей, так как дает общее представление о самом сорте. К данной группе относятся следующие характеристики:

– *фото* – текстовый тип данных, для уменьшения объема информации, в этой ячейке будет храниться ссылка на местоположение фотографии сорта, а не сама фотография;

– *направление использования* – тип текстовый («столовый», «технический», «универсальный»);

– *сила роста* – тип текстовый («сильнорослый», «среднерослый», «слаборослый»);

– *стабильность плодоношения* – тип текстовый («стабильный», «нестабильный»);

– *масса грозди* – тип числовой;

– *форма ягоды* – тип текстовый;

– *цвет ягоды* – тип текстовый;

– *размер ягоды, мм* – тип числовой;

– *окрас* – тип текстовый;

– *семена* – тип текстовый («семянной», «бесемянной»);

– *масса ягоды минимальная, максимальная, средняя многолетняя* – тип числовой;

– *индекс продуктивности* – тип числовой;

– *средний урожай с куста* – тип числовой.

Правовая основа – интересна сельхозтоваропроизводителям и селекционерам, информация о происхождении сорта (в том числе генетическом), годе его создания и внесения в Государственный реестр. Содержит следующие параметры:

– *генетическое происхождение* – информация о родителях (сортах, использованных для создания данного сорта), тип текстовый;

– *происхождение (вид)* – тип текстовый;

– *год создания* – тип числовой;

– *год внесения в реестр* – тип числовой;

– *автор(ы)* – текстовый тип;

– *патент* – текстовый тип;

– *обладатель прав* – тип текстовый, (автор сорта и обладатель прав на этот сорт).

Устойчивость (или восприимчивость) к болезням и вредителям – информация для виноградарей - производителей, генетиков. Содержит информацию об устойчивости/восприимчивости сорта к различным факторам (болезням, вредителям, лимитирующим факторам среды). Включает следующие параметры – устойчивость/восприимчивость к:

– *милдью* – тип числовой;

– *оидиуму* – тип числовой;

– *серой гнили* – тип числовой;

– *черной гнили* – тип числовой;

– *черной пятнистости* – тип числовой;

– *антракнозу* – тип числовой;

– *вирусным заболеваниям* – тип числовой;

– *филлоксере* – тип числовой;

– *хлорозу* – тип числовой;

– *известки* – тип числовой;

– *низким температурам* – тип числовой (для всех параметров, кроме устойчивости к низким температурам указывается балл устойчивости/восприимчивости, для последнего параметра – минимально допустимая температура).

Фенологические показатели – используется сельхозтоваропроизводителями для создания сортового конвейера, генетиками для подбора сортов для скрещивания, виноделами для выявления периода сбора урожая для планирования производства. Содержит следующие данные:

- *продолжительность вегетации*, тип числовой, указывается количество дней;
- *сокодвижение* – дата (число.месяц);
- *рост побегов и соцветий* – дата (число.месяц);
- *цветение* – дата (число.месяц);
- *рост ягод* – дата (число.месяц);
- *начало созревания* – дата (число.месяц);
- *полная зрелость* – дата (число.месяц).

Физико-химический состав – группа интересна виноградарям, виноделам, генетикам. Включает информацию о содержании в ягоде следующих веществ:

- *витамин В9* – тип данных числовой;
- *витамин С* – тип данных числовой;
- *катехины* – тип данных числовой;
- *антоцианы* – тип данных числовой;
- *лейкоантоцианы* – тип данных числовой;
- *сахар* – тип данных числовой;
- *кислота* – тип данных числовой;
- *глюкоацидометрический показатель* – тип данных числовой.

Дегустационные оценки – в основном интересна для виноделов, реже, аграриев. Содержит характеристики:

- *органолептические показатели* – описание вкуса ягод, тип текстовый;
- *дегустационная оценка винограда* – тип числовой;
- *дегустационная оценка изюма* – тип числовой;
- *дегустационная оценка сока* – тип числовой (указываются дегустационные баллы).

Эти показатели описывают столовые сорта.

Для технических сортов это показатели – дегустационная оценка вина:

- *сухого молодого* – тип числовой;
- *сухого без выдержки* – тип числовой;
- *сухого выдержанного* – тип числовой;
- *сухого марочного* – тип числовой;
- *сухого коллекционного* – тип числовой;
- *десертного молодого* – тип числовой;
- *десертного без выдержки* – тип числовой;
- *десертного выдержанного* – тип числовой;
- *десертного марочного* – тип числовой;
- *десертного коллекционного* – тип числовой (указываются дегустационные баллы).

Донор – используется селекционерами и генетиками, информация о том, может ли сорт быть донором устойчивости к тем или иным болезням, вредителям, лимитирующим факторам при скрещивании. Включает информацию:

- донор устойчивости к биотическим факторам: *милдью, оидиуму, серой гнили, черной гнили, черной пятнистости, антракнозу, вирусным заболеваниями, филлоксере, хлорозу, извести* – тип числовой (указывается балл);
- донор устойчивости к абиотическим факторам, в том числе: *к низким температурам* – тип числовой, указывается минимальная температура;
- *донор крупной ягоды* – тип числовой («да», «нет»);
- *донор мускатного аромата* – тип числовой («да», «нет»);
- *донор сроков созревания* – тип текстовый («да», «нет»);

- *дата созревания* – тип дата;
- *донор уровня урожайности* – тип текстовый («да», «нет»);
- *донор качества урожая* – тип текстовый («да», «нет»);
- *донор наличия семян* – тип текстовый («да», «нет»).

Однако для определения структуры БД словаря недостаточно, так как он не описывает, как связаны элементы, для этого строится логическая структура.

Этап 2. Логическая структура

Данный этап НИОКР предусматривает описание организации данных в таблицах и определение связи между этими таблицами. Результат должен быть представлен в виде диаграммы «сущность-связь».

Перед построением такой диаграммы нужно определить первичные ключи для всех спроектированных таблиц.

Ввиду того, что числовые ключи более удобны и их быстрее обрабатывает СУБД, по аналогии с базой о сортах, во все таблицы добавлены кодовые поля:

- таблица «Информация об областях» - добавлен атрибут *ID_Области*;
- таблица «Информация о почвах» - *ID_Почвы*;
- таблица «Информация о хозяйствах» - *ID_Хозяйства*.

Ключи – состоящие из одного атрибута – называются *простыми*. Однако не для всех таблиц можно определить простые ключи. Например, для таблицы «Продуктивность и качество» ключ будет состоять их двух атрибутов: *ID_Сорта* и *ID_Хозяйства*, так как один и тот же сорт может возделываться в нескольких хозяйствах, аналогично в хозяйстве могут возделываться несколько сортов. Поэтому только сцепка двух этих атрибутов дает уникальную строку в таблице. Такой ключ называется *составным ключом*.

Для связи почв с областями возделывания и хозяйствами используются таблицы, состоящие из двух атрибутов-ключей:

- таблица «Почвы областей» - *ID_Области, ID_Почвы*;
- таблица «Почвы хозяйств» - *ID_Хозяйства, ID_Почвы*;

На данном этапе построена схема данных, наглядно представляющая взаимосвязи между таблицами (рис).

Этап 3. Физическая структура

Физическое проектирование базы данных – создание схемы базы данных для конкретной СУБД, создание описания СУБД. Специфика конкретной СУБД может включать в себя ограничения на именованное объектов базы данных, ограничения на поддерживаемые типы данных и т.п.

Создавая таблицу базы данных необходимо определить тип данных для каждого столбца (было сделано на первом этапе). Помимо типов данных РСУБД позволяет ограничить возможные для ввода данные. Например, ограничить длину или принудительно указать на уникальность значения записей в данном столбце.

Эти ограничения дают контроль над целостностью данных и предотвращают:

- ввод текста в поле, в котором должно быть число;
- ввод слишком большого или маленького числа;
- создание сортов с одним и тем же именем.

Результатом проведения исследований является разработка структуры базы данных хозяйственно-ценных признаков сортов винограда. Информация, содержащаяся в базе, наиболее полно описывает сорт и не содержит лишней (маловажной) информации.

Кроме того, после проектирования словаря данных и разделения характеристик сортов на функциональные группы, разработаны макеты форм для вывода данных о сортах, указанные на «Главной странице» разработанного в разделе 3 интерфейса (рис.).

Общая характеристика сорта

Сорт	<input type="text"/>	Фото	<input type="text"/>
Направление использования	<input type="text"/>		
Сила роста	<input type="text"/>		
Стабильность плодоношения	<input type="text"/>		
Масса грозди, гр	<input type="text"/>		
Цвет ягоды	<input type="text"/>		
Форма ягоды	<input type="text"/>	Размер ягоды, мм	<input type="text"/>
Окрас	<input type="text"/>	Семена	<input type="text"/>
Масса ягоды, гр:			
минимальная	<input type="text"/>	средняя многолетняя	<input type="text"/>
		максимальная	<input type="text"/>
индекс продуктивности	<input type="text"/>	средний урожай с куста, кг	<input type="text"/>

Рис. Форма «Общая характеристика сорта»

Таким образом, в результате исследований проведен системный анализ агробиологических и физиолого-биохимических признаков сортов винограда разных групп по эколого-географическому происхождению: западноевропейские (*Convar occidentalis Negr.*), восточные (*Convar orientalis Negr.*), побережья Черного моря (*Convar pontica Negr.*), внутривидовые *V. vinifera L.* и межвидовые гибриды. Описание сортов сгруппировано по следующим признакам: общая характеристика сортов винограда; фенологические показатели сортов винограда; восприимчивость сортов винограда к вредным организмам; физико-химический состав сока ягод сортов винограда; органолептическая оценка сортов винограда и вин; сорта винограда – доноры хозяйственно-ценных признаков для использования в селекции; правовая основа технических сортов винограда.

Выводы. В результате проведенных исследований разработана структура базы данных хозяйственно-ценных признаков сортов винограда: проведен анализ требований, построен словарь данных (на основе анализа параметров, описывающих сорта, и выявления параметров, наиболее полно характеризующих сорт), разработаны логическая и физическая структуры базы. Разработаны формы предоставления информации о сортах из базы по различным функциональным группам.

Литература

1. Виноградарство; под ред. Раджабова А.К. – М.: ФГБНУ «Росинформагротех», 2017. 500 с.
2. Егоров Е.А., Воробьева Т.Н., Ветер Ю.А. Продуктивный потенциал промышленных виноградников // Аграрная наука. 2007. №1. С. 18-21.
3. Егоров Е.А., Ильина И.А., Петров В.С., Панкин М.И., Ильницкая Е.Т., Талаш А.И., Лукьянов А.А., Лукьянова А.А., Коваленко А.Г., Большаков В.А., др. Анапская ампелографическая коллекция (биологические растительные ресурсы): монография. Краснодар: ФГБНУ Северо-Кавказский федеральный научный центр садоводства, виноградарства, виноделия, 2018. 194 с. <https://elibrary.ru/item.asp?id=36539666>
4. Веденяпин, Г.В. Общая методика экспериментального исследования и обработки опытных данных. Третье издание, дополненное и переработанное. – М.: «Колос», 1973. 199 с.
5. Вольф, В.Г. Статистическая обработка опытных данных. – М.: «Колос», 1966. 259 с.
6. Скотт В. Эмблер, Прамодкумар Дж. Садаладж. Рефакторинг баз данных: эволюционное проектирование. — М.: «Вильямс», 2007. С. 368.
7. Томас Коннолли, Каролин Бегг. Базы данных. Проектирование, реализация и сопровождение. Теория и практика – 3-е изд. – М.: «Вильямс», 2003. С. 1436.
8. К. Дж. Дейт Введение в системы баз данных – 8-е изд. – М.: «Вильямс», 2006. С. 1328.
9. Голицина О.Л., Максимов Н.В., Попов И.И. Базы данных: Учебное пособие. – М.: ФОРУМ: ИНФРА-М, 2003. С.352.