

УДК 631.4

КОНЦЕПЦИЯ ГЕНЕРАТОРА (МАШИНЫ) КЛАССИФИКАЦИЙ ПОЧВ

© 2016 г. В. А. Рожков

*Почвенный институт им. В.В. Докучаева,
Россия, 119017, Москва, Пыжевский пер., 7, стр. 2
e-mail: rva39@mail.ru*

Классификации почв, соответствующие современному уровню информационных технологий, способны обеспечить необходимую прозрачность и доказательность их концепций, применение формальных методов для количественных оценок и критериев качества суждений. Для мирового почвоведения сейчас характерно отсутствие серьезных исследований по теории классификации, которая подменяется методами анализа данных. Как попытка восполнить этот пробел в статье предлагается новый подход к построению исчерпывающей (comprehensive) классификации на основе идеи классификации-перечисления (по Ю.А. Воронину). Цель классификации определяется через фиксированную систему выбранных почвенных показателей (признаков), организованных в виде графа классификации-перечисления. Число объектов в классификации определяется числом выбранных признаков и градаций их значений. Грани графа отражают описания представительных образов (типов) классификационных таксонов. По сходству с ними происходит распознавание новых почв. Впервые классификация получила аналитическое выражение, и ее формула позволяет кодировать числом вектор-описаний почв и, наоборот, восстанавливать описание по коду – своеобразная компактная база данных. В формирующееся классификационное пространство могут помещаться другие классификации и информационные материалы. Для информационного обеспечения и реализации обсуждаемых подходов создан специальный сайт www.infooil.ru.

Ключевые слова: классификация-перечисление, представительный профиль, типовой образ, голотип, предмет классификации.

DOI: 10.19047/0136-1694-2016-85-115-130

ВВЕДЕНИЕ

Классификация – это способ познания объективной реальности, предполагающий процесс разработки, его результат и процедуры использования – определение принадлежности (распознавания) предмета или ситуаций к известным их классам. Суще-

ствующие классификации почв отражают опыт, знания и пристрастия отдельных коллективов авторов. Отсутствие общепринятых классификаций почв в стране и мире, очевидно, связано с целым рядом причин. Практически все существующие классификации почв построены на использовании эмпирического материала, содержащего отражение пристрастий автора, находящегося в плену уже сложившихся представлений о готовой классификации¹. Однако субъективные суждения не могут быть *очевидными* для всех и требуют формального выражения в “алгоритме очевидности”. Согласно ему, компьютер выполняет доказательство, а не “угадывание очевидной вещи”, для чего ему нужно “по готовой конструкции проверять, удовлетворяет ли она тем или иным свойствам” ([Управление...](#), 1976, с. 206–207).

Даже среди авторитетных ученых преобладает некорректное понимание самого понятия “классификация” – от нее требуют сразу многих иногда противоречивых возможностей: отражения генезиса, плодородия, перспектив эволюции, экономических показателей и др.

Цели чаще всего декларативны и не обосновываются обоснованием комплекса используемых свойств. Индексы генетических горизонтов – явно недостаточные диагностические признаки, они в такой же степени, как и почвы, требуют обоснования конкретными показателями свойств и состава. Часто нет четких критериев качества построенных классификаций. На низком уровне находится формализация всех этапов решений, отсутствие количественных показателей информативности признаков, деления таксонов и способов распознавания новых почв.

Для мирового почвоведения характерно отсутствие серьезных исследований по *теории классификации*, которая подменяется *методами* анализа данных. В подтверждение можно привести содержание одного из специальных журналов в этой области ([Advances in Data Analysis and Classification, 2016](#)):

- A principal component method to impute missing values for mixed data.
- Extreme logistic regression.

¹Примечательно, что практически все учебники почвоведения, начиная с первого ([Белецкий, 1893](#)), имеют практически одинаковую структуру глав и параграфов ([Добровольский, Рожков, 2009](#)).

- A multilevel finite mixture item response model to cluster examinees and schools.
- A comparison of reliability coefficients for psychometric tests that consist of two parts.
- Supervised clustering of variables.
- [The k-means algorithm for 3D shapes with an application to apparel design](#)

Следует отметить, что часто бывает сложно разделить что относится к теории, а что к методам. Например, рН и влажность можно интерпретировать как свойства, так и количество.

В ряде предшествующих статей сделана попытка исправить это положение ([Рождков, 2012а, 2012б, 2014](#); [Шрейдер, 1983](#)). Как продолжение работы в настоящей статье предлагается новый подход к построению *исчерпывающей* (comprehensive) классификации на основе идеи *классификации-перечисление* ([Воронин, 1970](#)). Идея возникла при разработке теории классификации в геологии. Исчерпывающей ее следует называть потому, что она отражает однозначную связь между признаками и объектами. И количество значений признаков определяет число описываемых объектов.

В 1936 г. А. Тьюрингом предложен абстрактный универсальный исполнитель – логическая вычислительная конструкция. На ней можно имитировать операции, которые выполняют реальные вычислительные машины. Эта конструкция названа машиной Тьюринга (<http://www.infl.info/turing>). Употребляются также понятия: “машина желаний”, “мегамашина” ([Новейший философский словарь, 2001](#), с. 609–613). По аналогии с этим, возникла идея создать машину или генератор целевых классификаций, которые, используя фиксированную совокупность признаков объектов, соответствующих намеченной цели исследования, создают классификационные структуры.

ТЕОРИЯ ВОПРОСА

Цель классификации отражается через фиксированную систему учтенных показателей (признаков), принимаемых за основания деления. Под признаком понимается пара, состоящая из наименования показателя, характеризующего объект, и его значения. Значения признаков могут задаваться в различных *шкалах*, определяющих допустимые виды их обработки математическими

методами ([Рожков, 2011](#)). Понятие признака обобщает конкретные наименования и значения свойств, состава и отношений объекта.

Точнее следует говорить о *предмете* – части объекта, в данном случае почвы. Предмет – это срез, проекция общего понятия почвы, которая может быть представлена с разных сторон или точек зрения, разными предметами, отражающими плодородие, генетические или мелиоративные аспекты, экологические функции и т.д. Всякий раз формируется новый предмет, представляемый совокупностью соответствующих почвенных показателей – состава, свойств, режимов, условий почвообразования, функций почвы. Изменение состава учтенных почвенных признаков равносильно созданию новой классификации.

Предмет исследования не может быть задан иначе, как через совокупность признаков – описание. Полезно повторить несколько соображений от [Дж.Ст. Милля \(1914\)](#), более полно изложенных ранее ([Рожков, 2012а](#)).

“Нет ни одного свойства предметов, которого нельзя было бы, если угодно, принять за основание для классификации, ...и мы скорее всего выберем для этой цели свойства простые, легко представимые и заметные с первого взгляда...” ([Милль, 1914](#), с. 644). “...предметы следовало бы классифицировать по возможности на основании таких свойств, которые служат причинами многих других или, ...такие, которые... сами представляли бы собою такие свойства, на которых нам было бы всего полезнее сосредоточить наше внимание” (с. 645).

Таким образом, должна формироваться соответствующая система показателей – *информационная база классификации* ([Шишов и др., 1985](#); [Рожков, Столбовой, 1988](#); [Рожков, 2005](#)). Показатели должны быть общими для всех почв “без пропуска”. Значения *всех* выбранных показателей должны быть указаны для *каждой* почвы, и в расчет принимаются все возможные комбинации свойств и их значений. Эти комбинации определяют *классификационное поле (пространство) и число объектов* в нем.

Например, пусть условно выбраны градации степени гумусированности: 1 – слабо-, 2 – средне- и 3 – высокогумусированный; 1 – насыщенный, 2 – ненасыщенный основаниями. Классификация перечисление возможных объектов будет иметь 18 объектов (рис. 1), и никаких других вариантов (сочетаний градаций кислот-

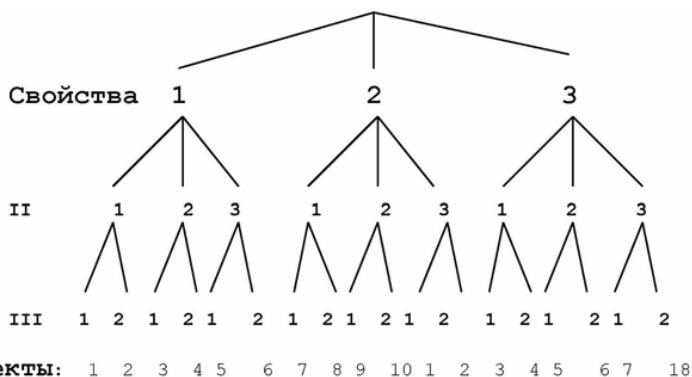


Рис. 1. Классификация-перечисление.

ность: 1 – низкая, 2 – средняя, 3 – высокая и третий признак: свойств) в данной классификации не может быть. Наоборот, могут встречаться несуществующие в природе сочетания, например, насыщенность сильно кислых почв. Такая избыточность не мешает организации и анализу данных, но обеспечивает сбалансированность классификации.

Очевидно, что перестановка свойств по уровням не меняет числа описываемых ими объектов. В то же время можно видеть, как изменяется классификация от включения нового признака: Например, если на рисунке добавить признак с тремя значениями: слабо-, средне-, сильно-, то вместо 18 в структуре будет уже 54 объекта. Но это будет уже другая классификация.

Классификация-перечисление выполняет следующие функции:

- формально фиксирует объем и содержание множества классифицируемых объектов по совокупности учтенных свойств;
- дает основу для выработки и проверки понятий о множестве этих объектов;
- служит средством эффективного кодирования эмпирических данных, исключая содержание специальной базы данных;
- позволяет ввести сразу всю систему имен в этом множестве, т.е. всегда является исчерпывающей системой;
- обеспечивает возможности для оптимизации эмпирических исследований;

– открывает возможности для арифметизации многих действий.

Число объектов в классификации, равное числу ветвей полученного графа, определяется как произведение чисел градаций значений признаков:

$$N = \prod_{j=1}^m G_j$$

где m – число учтенных признаков. Здесь $N = 3 \cdot 3 \cdot 2 = 18$.

Существует простая связь между номером объекта и его описанием, которая позволяет заменить одно другим. Переход от вектора свойств к номеру осуществляется согласно выражению (Рожков, 1989):

$$N = Z [G_j(Q_{j-1} - 1) + K_j], \text{ при } Q_1 = K_1,$$

где $j \geq 2$, где Z условное обозначение рекуррентного соотношения; K_j – градация j -ого признака; Q_j – рабочий массив.

В том случае, когда число свойств невелико, приведенная формула может быть заменена многочленом, еще более удобным в использовании:

$$N = \sum_{j=1}^{m-1} K_j \prod_{j+1}^m G_j + K - \sum_{j=1}^{m-1} \prod_{j+1}^m G_j$$

Приведенные уравнения, по существу, представляют собой *аналитическое выражение* классификации перечисления². В перспективе это дает возможность размещать такую систему на мало габаритных гаджетах и работать в режиме off line.

Ветви дерева на рис. 1 представляют собой описания именно тех *представительных (базовых, типичных или просто типов) образов*, на которые обычно ссылаются почвоведы. Таким образом, снимается неопределенность часто употребляемого понятия *типичного* разреза (профиля), поскольку тогда неясно, это экспертное, усредненное, модальное, медианное описание или *голотип* – элемент таксона, наиболее сходный со всеми другими эле-

²Составлена компьютерная программа, которая по перечню градаций значений признаков в классификаторах присваивает почве номер в классификации перечисления, и наоборот – по номеру раскрывает все описание почвы (Рожков, 1989). На пример, 6-я почва имеет описание 1-3-2, а 18-я – 3-3-2.

ментами этого таксона. Новые почвы, описанные теми же признаками, распознаются по сходству с представительными или по вычисляемому номеру на графе.

Очевидно, что такой же граф можно строить по любым другим основаниям: по составу генетических горизонтов, элементарным почвенным процессам, по собственным свойствам почвы и т.д. Это будут различные по цели классификации, построенные на разных основаниях, поскольку именно признаки, ставшие основаниями деления, отражают цель классификации.

Таким образом, реализуется самый простой и корректный способ создания любой классификации. Здесь речь идет о порядковой или номинальной шкале признаков. Более сильные шкалы потребуют преобразования (квантования) в такие же более слабые шкалы.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В предлагаемом подходе четко формулируются принципы создания классификации, применимые к созданию новых классификаций, как это требовалось в работе ([Зольников, 1955](#)). По существу, они более всего отвечают требованию определенности взаимосвязи состава признаков и числа всех возможных объектов. Ни одна из почвенных классификаций не обладает такой определенностью относительно количества выделяемых почв и формы их представления. Обычно этот вопрос решается обратным путем: если известные почвы не укладываются, в их классификацию корректируют исходные или добавляют какое-либо новое основание, которое будет отделять эту почву от других, не заботясь о роли этого основания для разделения других почв. Обсуждаемый подход ограничивает произвол способов группировки, делает прозрачным принцип выбора оснований деления. Выбор признаков и соответствие их целям классификации может проверяться количественными методами, реальными наблюдениями.

Источниками систематизированных списков почвенных признаков и методов их определения могут служить материалы к почвенным информационным системам ([Автоматизированная..., 1980](#); [Методическое руководство..., 1986](#); [Колесникова и др., 2010](#)). В зависимости от целей будут выбираться разные наборы признаков (информационная база классификации), т.е. появится

много классификаций. Почвенные показатели могут быть разгруппированы соответственно видам классификации: субстантивной, факторной и др. Дискуссии по поводу существа группировки сведутся к обсуждению состава признаков, т.е. сугубо профессиональным вопросам, которые более существенно обоснованы экспериментами, чем интуитивные представления.

Простота алгоритма и его реализации создает перспективы систематизации самих классификаций для последующей их комбинации в многоцелевые агломерации. Все эти варианты будут исчерпывающими, узкоцелевыми, интенциональными, поскольку в рамках выбранного классификационного поля они охватывают все множество возможных в нем объектов. Такие классификации необходимы для корректной постановки и решения задач построения разнообразных других классификаций.

Содержательные примеры описанного подхода приведены в предыдущих публикациях по теории классификации почв. Особенно наглядно выглядит фрагмент факторной классификации ([Иванова, Розов, 1960](#)). В графической развертке он представлен на рис. 2.

В приведенной работе применен формальный подход к построению классификации: назначены четыре основания деления (почвообразование, классы, подклассы, режимы), даны их значения ($G[4] = 4, 6, 3, 3$), указаны все возможные при этом факторно-географические “ниши” условий почвообразования ($N = 4 \cdot 6 \cdot 3 \cdot 3 = 216$), часть из которых была заполнена известными типами почв. Таким образом, в этой классификации имеется место для 216 факторно-обусловленных природных ситуаций (экологических ниш) для почв. Не все из них нашли свое соответствие с известными почвами – пропуски отмечены знаком вопроса. Такая ситуация подвигает исследователя на восстановление пропусков и выделение новых типов почв (как в таблице Менделеева!).

Одновременно оказалось осуществимым высказанное требование Б.Г. Розанова, чтобы классификация указывала пути к открытию новых типов почв. Более того, анализируя географию почв и соответствующие им природные условия, взятые за основания деления, явно представленные на приведенной структуре, можно делать прогнозы изменения почв или их эволюции в связи с глобальными изменениями климата.

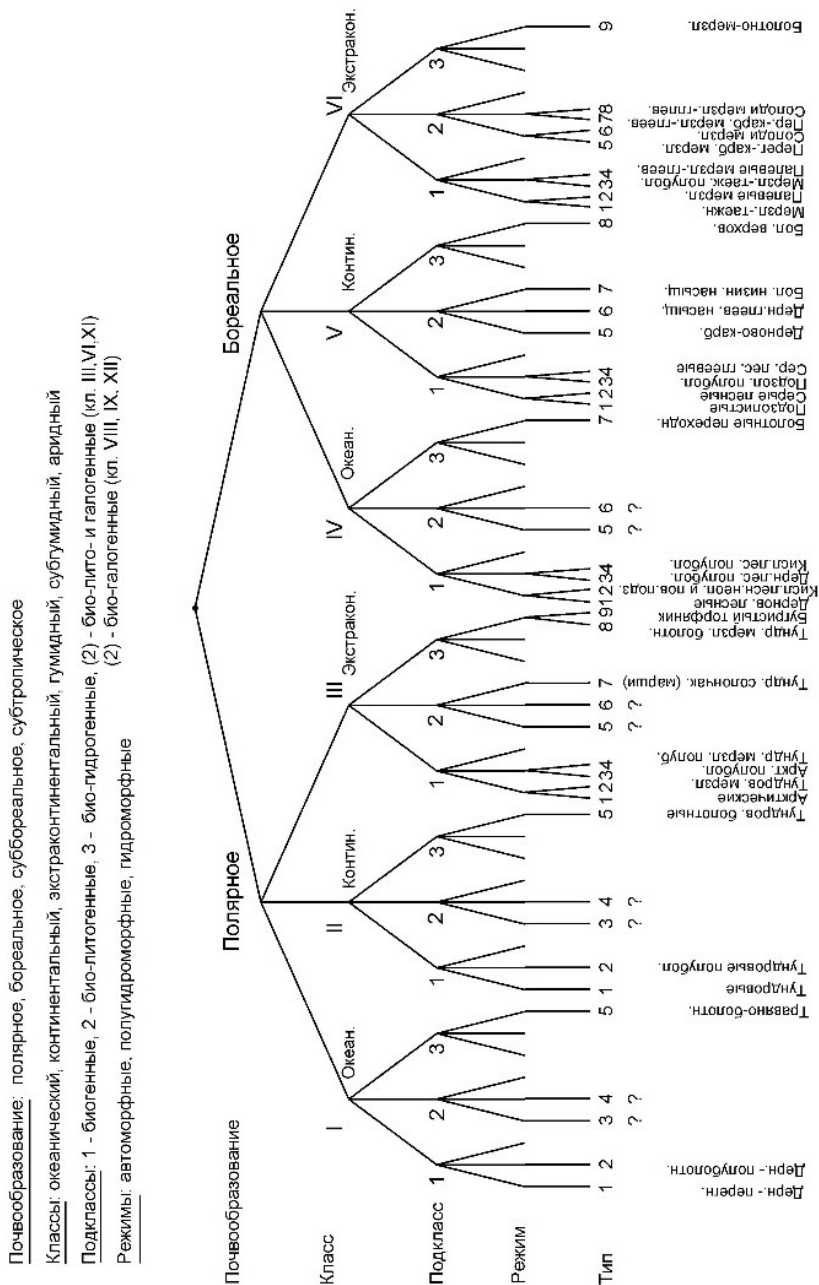


Рис. 2. Фрагмент факторной классификации.

В классификационное поле можно “подвесить” любую другую классификацию, лишь бы ее объекты (их предметы) содержали показатели, принятые в классификации-перечислении. Создание и использование таких перечислений по генезису, факторам, плодородию, процессам, режимам, разным комплексам свойств и т.п. делается вполне оперативно. Это позволит идентифицировать разные почвенные предметы. Таким образом, будет решена проблема смены со временем систем классификации или легенды почвенной карты без существенных исследований корреляций между их вариантами.

Фундаментальность описанного принципа трудно переоценить, так как он несет принципиально новые идеи о структуре и логике создаваемых классификаций, а также реальное подтверждение того, что цель классификации отражается через систему учтенных показателей, которые не должны эклектично смешиваться, поскольку это не приведет к многоцелевой классификации, а только увеличит ее неопределенность.

Классификация-перечисление – соответствует *исчерпывающей* классификации объектов, заданных конечной (фиксированной) совокупностью признаков, соответствующих намеченной цели группировки, в пространстве значений которых представлены все классифицируемые объекты и будут распознаваться вновь представляемые описания почв. Представление о классификации-перечислении может распространяться и на картографические объекты (рис. 3).

Нижняя карта уникальных комбинаций контуров трех верхних карт представляет собой аналог классификации-перечисления. Предельное число новых контуров, равное здесь 24, явно больше реальных пересечений карт, однако тут важна аналогия и перспектива формализации процедуры оверлея карт.

Как отмечалось выше, наиболее часто трудности создания классификаций почв возникают из-за некорректной постановки задач, пытающихся совместить в единой классификации происхождение, строение, состав, плодородия и одновременно отражать роль общественного производства. Очевидно, что подобное смешение целей классификации некорректно, но для картографического варианта вполне реальная ситуация.

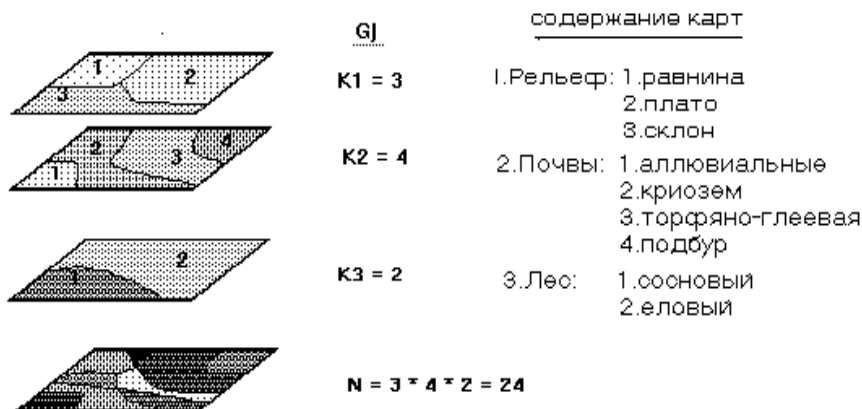


Рис. 3. Классификация-перечисление на картографических объектах.

Состав учетных признаков должен соответствовать цели классификации, но обеспечить это не просто. Например, состав генетических горизонтов в профиле бесспорно генетический признак. Однако индексы генетических горизонтов явно недостаточно четко отражают генетические особенности почв – они в такой же степени, как и почвы, требуют своего обоснования конкретными показателями свойств, состава, режимов. В прежних классификациях индекс А1 выделяется практически во всех почвах, следовательно, он не служит разделению почв, т.е. он не имеет информативности при идентификации типов. И другие горизонты, одинакового индекса не всегда являются сходными по реальным свойствам разных почв. Кластер-анализ данных профильных кодов (в отличие от процессных) [И.П. Герасимов \(1976\)](#) подтверждает слабую информативность индексов горизонтов ([Рожков, Скворцова, 2009](#); [Рожков, 2012а](#)).

Классификация может выделять непересекающиеся (не имеющие общих объектов) классы эквивалентности или нечеткие (размытые) таксоны. В данном случае речь идет о классификации, как системе классов эквивалентности. Отношение эквивалентности обладает свойствами рефлексивности (xRx), симметричности ($xRy \rightarrow yRx$) и транзитивности ($xRy \& yRz \rightarrow xRz$), где R – некоторое отношение – сходства, различия, подобия и др. Формально классификация представляется множеством объектов A , которое

делится на классы: $A = A_j$, где $j = 1, 2, \dots, k$ и k – число классов; $A_j \neq \emptyset$, т.е. все классы непустые (содержат хотя бы один объект); $A_j \cap A_l = \emptyset$, где $i, j = 1, 2, \dots, k$ и $j \neq l$, т.е. классы не пересекаются; $A_j = A$ – объединение классов составляет исходное множество. Однако в перспективе более эффективными должны оказаться классификации на принципах нечетких (размытых) множествах.

В тексте [Классификации... \(1977\)](#) видно, как авторы, работавшие в разных природных зонах, детально описывают “свою” почву и приводят их отличия от “соседей” по разным показателям. Иными словами, каждый исследователь имел собственное представление о некотором *type* почвы. Речь идет о *type* не в смысле *типа почвы*, а *типичного* представителя некоторого класса или представительном образе объектов, соответствующего ветви графа классификации-перечисления.

Тип, широко используемый термин в классификации, [Дж.Ст. Милль \(1914\)](#) определял как образец класса объектов, обладающий признаками этого класса, выраженными в наиболее резкой степени – классом руководит не определение, а тип. Именно тип может быть положен в основу определения таксонов классификации: “Естественные группы устанавливаются ... типом, а не определением” (с. 648).

[Ю.А. Шрейдер \(1983\)](#) рассматривает понятие типа в двух аспектах: первый обозначает конкретное описание одного из объектов класса выбранного в качестве эталона или модели для остальных, а второй – прообраз, основную форму, допускающую отклонения, что в некотором смысле соответствует понятию архетипа. “Архетип позволяет представить объекты путем описания, эксплицитно указывающего на общее и одновременно особенное объектов данного типа. Бывает невозможно задать архетип через один типичный образец, тогда его рассматривают как образ (гештальт), заданный многообразием типичных образцов...” (с. 93–94).

Практика показывает, что типичный профиль (разрез) чаще всего выделяется экспертно-эмоционально в острых и длительных дискуссиях (“Мы сделали 20 разрезов, пока нашли типичный профиль!”). Тогда трудно определить, что получается – реальный образ или архетип. Видимо, в большей мере к определению подходит архетип. Статистически определенный образ – средний, меди-

анный, модальный – ближе к реальному объекту класса. Голотип – это конкретный объект класса, наиболее сходный с другими представителями таксона. Классификация-перечисление целиком состоит из типов – реальных *образов* классов. Нужно отметить, что граф классификации-перечисления не представляет иерархию в строгом смысле: признаки могут переставляться на любые уровни, а узлы (значения признаков) не обладают свойством целостности (холархии) ([Пригожин, Стенгерс, 2005](#)).

Предпосылки приложения концепции классификации-перечисления начаты в разработке информационно-советующей системы МЕРОН ([Рожков, 2011](#)). Тогда было установлено, что ограничением ее использования станет количество выбранных признаков и их значений, от которого зависит разрядность (величина) кодов почв. Это не проблема для компьютеров, допускающих любой размер целого числа (например, ЭВМ МИР), в других случаях, возможно, следует использовать методы нумерационного кодирования ([Амелькин, 1986](#)) или программировать процедуру конвертации кода по мере его вычисления в текстовый формат (и обратно).

Реализация описанных подходов к генерации классификаций почв требует качественной и хорошо организованной информационной основы.

ЗАКЛЮЧЕНИЕ

Создается формализованная система (“машина”) генерации любых тематических классификаций перечисления. В качестве исходной информации для этого используется совокупность почвенных признаков, соответствующих цели задуманной классификации. Реляционные отношения классификаторов значений выбранных признаков формируют систему представительных типов таксонов классификации. [А.Л. Субботин \(2001\)](#), справедливо отмечал, что требование создать “окончательную”, “абсолютно совершенную” классификацию, за исключением, быть может, простейших случаев, выглядит несерьезно; оно способно лишь дезориентировать в оценке реальных шагов в классификационной работе и породить неконструктивный скептицизм. Очевидно, что классификация-перечисление не является той классификацией почв, надежды и споры о которой, не прекращаются в мировой

науке. Не пересекаются ее функции и с WRB. Тем не менее, она послужит эффективным операционным инструментом генерации целевых классификаций, исчерпывающих в том плане, что они описывают все возможные типы объектов, возможных в заданном пространстве признаков.

Следуя рассуждению [Дж.Ст. Милля \(1914\)](#) о роли научных классификаций, про обсуждаемую классификацию перечисление можно сказать, что эти принципы приложимы ко всем случаям, в которых человечеству приходится умственно координировать различные части того или другого обширного предмета. Принципы эти столь же полезны в классификациях, составляемых в целях того или другого искусства или практической жизни, как и в тех, которые должны удовлетворять целям науки

Реализация описанных подходов к генерации классификаций почв требует качественной и хорошо организованной информационной основы.

СПИСОК ЛИТЕРАТУРЫ

1. Автоматизированная информационно-поисковая система ПОЧВА. М.: Почв. ин-т им. В.В. Докучаева, 1980. Ч. 1–5.
2. *Амелькин В.А.* Методы нумерационного кодирования. Новосибирск: Наука, 1986. 160 с.
3. *Белецкий Н.* Почвоведение. Образование почвы, ее состав и свойства. Виды почв, их классификация, бонитировка и картография. М.: Типо-литография Т-ва И.Н. Кушнеревъ и К^о, 1895. 476 с.
4. *Воронин Ю.А.* Геология и математика. Новосибирск: Наука, 1970. 224 с.
5. *Герасимов И.П.* Элементарные почвенные процессы как основа для генетической диагностики почв // Генетические, географические и исторические проблемы современного почвоведения. М.: Наука, 1976. С. 193–208.
6. *Добровольский Г.В., Рожков В.А.* [Один из первых отечественных учебников почвоведения](#) // Почвоведение. 2009. № 11. С. 1404–1407.
7. *Зольников В.Г.* Об основных методологических принципах генетической классификации почв // Почвоведение. 1955. № 11. С. 70–79.
8. *Иванова Е.Н., Розов Н.Н.* Классификация почв СССР // Докл. советск. почвоведов к VII Междун. конгр. в США. М.: Изд-во АН СССР, 1960. С. 280–293.
9. Классификация и диагностика почв СССР. М.: Колос, 1977. 224 с.
10. Методическое руководство по описанию почв в системе информационной базы классификации почв. М.: ВАСХНИЛ, 1986. 126 с.

11. Колесникова В.М., Алябина И.О., Молчанов Э.Н., Шоба С.А., Рожков В.А. [Почвенная атрибутивная база данных России](#) // Почвоведение. 2010. № 8. С. 899–908.
12. Милль Дж.Ст. Система логики силлогистической и индуктивной. М.: Издание Г.А. Лемана, 1914. 880 с.
13. Новейший философский словарь. Мн.: Интерпрессервис; Книжный дом, 2001. 1280 с.
14. Пригожин И., Стенгерс И. Время, Хаос, Квант. К решению парадокса времени. М.: КомКнига, 2005. 232 с.
15. Рожков В.А. Алгебра WRB (формализация концепции) // Тр. Всерос. конф. “Экспериментальная информация в почвоведении: теория и пути стандартизации”. М.: Изд-во Моск. ун-та, 2005. С. 73–82.
16. Рожков В.А. [Дуализм основных понятий классификации почв](#) // Почвоведение. 2014. № 1. С. 3–9.
17. Рожков В.А. [Классиология и классификация почв](#) // Почвоведение. 2012а. № 3. С. 259–269.
18. Рожков В.А. [Об информационном подходе в классификации почв](#) // Бюл. Почв. ин-та им. В.В. Докучаева, 2012б. Вып. 69. С. 4–23.
19. Рожков В.А. [Формальный аппарат классификации почв](#) // Почвоведение. 2011. № 12. С. 1411–1424.
20. Рожков В.А. [Экономный код почвенных данных](#) // Бюл. Почв. ин-та им. В.В. Докучаева. 1989. Вып. 53. С. 32–35.
21. Рожков В.А., Скворцова Е.Б. [Тектология почвенной мегасистемы \(общность организации и анализа данных\)](#) // Почвоведение. 2009. № 10. С. 1155–1164.
22. Рожков В.А., Столбовой В.С. Построение классификации почв СССР с использованием автоматизированной информационной системы // Тр. Почв. ин-та им. В.В. Докучаева. 1988. С. 99–108.
23. Субботин А.Л. Классификация. М., 2001. 97 с.
24. Управление, информация, интеллект. М.: Мысль, 1976. 383 с.
25. Шишов Л.Л., Рожков В.А., Столбовой В.С. Информационная база классификации почв // Почвоведение. 1985. № 9. С. 9–20.
26. Шрейдер Ю.А. Систематика, типология, классификация // Теория и методология биологических классификаций. М.: Наука, 1983. 169 с.
27. [Advances in Data Analysis and Classification](#). 2016. V. 10 (1).

THE CONCEPTION OF SOIL CLASSIFICATION GENERATOR (MACHINE)

V. A. Rozhkov

*V.V. Dokuchaev Soil Science Institute,
Russia, 119017, Moscow, Pyzhevskii 7, bld. 2
e-mail: rva39@mail.ru*

The classifications of soils, which correspond the actual level of informational technologies, are able to provide the necessary clarity and conclusiveness of their conceptions, the implementation of formal methods for qualitative assessments and criteria of quality of estimations. The world soil science is now characterized by the absence of fundamental investigations on theory of classification, which is replaced by methods of data analysis. This paper suggests a new approach to creation of comprehensive classification as an attempt to fill a gap in this area. This approach is based on the idea of allotment classification (by Yu. A. Voronon). The purpose of classification is determined by the fixed system of the selected soil indices (specificities), which are organized as a graph of allotment classification. The number of objects in classification is determined by the number of selected specificities and their value grades. The edges of the graph reflect the descriptions of meaningful modes (types) of classification taxons. The recognition of new soils is made on the basis of similarity to them. It is the first time when the classification obtained analytical expression. Its equation allows us to code the vector of description of soils by a number, and as contrary, to restore the description on the basis of a code. That looks like an original database. Other classification and information materials may be moved into the forming classification space. For the purposes of supporting and realization of discussed approaches we created the special internet page: www.infosoil.ru.

Key words: allotment classification, representative profile, typical mode, holo-type, the subject of classification.