**Инструкция**

**по работе с программным комплексом Petros** **3.2**

**Часть 1. Подготовка анализов и ввод данных**

Tomas G. Petrov, Sergey V. Moshkin

**Abstract**

Инструкция предназначена для ознакомления пользователя с начальными этапами работы с программой Petros3.2. Программа создана для реализации возможностей информационного языка-метода ***RHAT*** по качественно-количественному кодированию составов объектов любой природы с одновременным созданием баз данных. Метод позволяет выполнять два взаимно дополняющих класса работ: 1) линейное алфавитное упорядочение информации о составах с целями иерархического периодического группирования и поиска аналогов составов, 2) диаграммное ̶ на базе трёх информационных характеризаций и их связей с первичными данными ̶ представление случайных и упорядоченных во времени или пространстве совокупностей составов для изучения, в частности, двух фундаментальных процессов изменения составов ̶ смешения и разделения.

Рассчитываются, коэффициенты корреляции, расстояния, статистические данные. Обеспечен экспорт результатов в Word и Excel. Метод обсуживает весь количественный диапазон любого анализа, а потому программа предъявляет повышенные требования к качеству исходных данных. По умолчанию предполагается работа с химическими составами в форматах оксидов и/или химических элементов. Загружены три алфавита: химический, фрагментированный минеральный и кристаллохимический группы турмалина.

Описание метода, программы (общее) и полученные результаты представлены на сайте ResearchGate под именем Tomas G.Petrov.

На сайте автора метода http://tomaspetrov.ru представлены: метод, программа и иные, связанные прямо и косвенно с методом ***RHAT***, научные направления и результаты

Находящаяся на сайте программа Petros3 доступна для скачивания на безвозмездной основе.

 **Оглавление**

**0. Введение 1**

**1. Главное (рабочее) окно программы 2**

**2. Ввод ссылки на источник данных 5**

**3. Требования к массиву анализов 6**

**4. Формирование файла 8**

**5. Ввод файла 9**

**0. Введение**

Метод обработки химических анализов горных пород, для которого создавалась программа, возник и описан 1971 году как реакция на информационный взрыв в геологии в 50-60-х годах. Название метода " RHA" – аббревиатура названий составных компонентов метода: R – ранговая формула, H – информационная энтропия К.Шеннона, A – анэнтропия, предложенная Т.Г.Петровым (Петров 1971). Позднейшее развитие метода привело к появлению еще одного параметра Т ̶ толерантности (Петров 2007), отказу от использования символа "Н", совпадающего с символом водорода в химических анализах и замене его на "Е" (entropy). В полном названии метода исходные символы сохранились ̶ RНAT, в обозначениях же используемой версии программы и в её приводящемся ниже описании "Н" и "Е" используются как синонимы.

 Приводимые ниже рекомендации продиктованы как требованиями работы программы, так и опытом работы, который может быть не обязательно оптимален для другого пользователя. Существует учебное пособие по методу RHA, центральному для организации всей программы (но не работы с ней) Т.Г. Петров, О.И.Фарафонова Информационно-компонентный анализ. Метод RНA СПб. Изд-во ЛГУ. 2005. 168с.; общее описание программы: Т.Г. Петров, С.В. Мошкин, Метод RHA и его реализация в программном комплексе Petros-3. Вычисления в геологии. 2011, №1, С. 50-53; Petrov T.G., Moshkin S. V. Method RHAT and its implementation in the software package PETROS-3 News of science and educations 2 (26) 2015 pp. 70-80.

Основное внимание при создании программы было направлено на работу с объектами геологии со стороны химических и минеральных составов. Поэтому ниже приводимые примеры относятся к этим двум областям.

Поскольку структура программы определяет формат материала, подготавливаемого к вводу, поэтому начнём с обзора того, что ждёт пользователя, когда он её откроет.

**1. Главное окно программы**



Главное окно включает (сверху вниз): 1) строку меню, к ней мы будем обращаться по мере необходимости, 2) инструментальную панель, 3) рабочую область и 4) строку состояния (под рабочей областью).

**Идем по номерам кнопок НИЖНЕЙ** – **инструментальной панели**



Инструментальная панель содержит кнопки, обеспечивающие быстрый доступ к основным пунктам меню. В панель включены следующие кнопки (здесь они перенумерованы):

1 – Queries (клик) – запросы, по которым будут производиться подавляющее большинство действий с файлами. При нажатии спадает окно, содержащее перечень файлов, к которым ранее были обращения. Упорядочение файлов в формате txt

2 – Data Sets – окно-список файлов, имеющихся в базе; здесь же справки об источнике данных, количестве анализов в файле и принятом для файла стандарте длины ранговой формулы (***n***), номере алфавита: «0» – химический . Упорядочение файлов – числовое. Здесь выделенный файл можно удалить "‒". Внизу справа кнопка отправить сигнал на кнопку 1 – Queries – для начала работы с файлом.

3 – Создание "списков" анализов, составляемых из отдельных анализов, уже имеющихся в Базе данных (см. ниже).

4 – Библиография. Источнику данных присваивается номер по мере поступления (или иначе) книг, статей, WEB. Под одним библиографическим номером могут быть и один анализ и много файлов с сотнями анализов. НО! Нужно иметь в виду, что анализы из одного файла на диаграмме НА будут иметь одинаковые значки. Поэтому, формируя файл, надо подумать о его использовании в работе – он особая группа. Если библиографического описания нет – писать что-то вроде: «ССА» =«Сер собств ан» = серия собственных анализов.

5 – Ввод данных с их набором внутри программы.

6 – Ввод таблицы данных, подготовленной вне программы. Данные для расчетов формируются в Excel с описаниями, начала которых в некоторой степени регламентируются для обеспечения нормальной работы программы. Сформированная таблица по шагам, диктуемым программой, вводится с автоматической проверкой. Ниже приводится отдельная серия скринов.

7 – Редактирование уже введенных данных. Возможно изменение: названия объекта, таблицы, количества компонентов при расчёте (при стандартизации), исправление ошибки в числе, тексте описания, удаление анализа из таблицы исходных данных (Внимание!) с сохранением номеров остающихся анализов в исходной таблице, что важно при выдаче анализов на диаграммы.

8 – При выделенном наборе данных в п..1 – выдача исходных данных

9-15 – Варианты представления анализов (с/без Н2О, приведение к 100% и пр.). (Расчётные данные)

16 – «Любимая» диаграмма, открывающаяся по умолчанию. Её заранее выбирают по кнопке 20.

17 – Диаграмма XY – оси выбираются или назначаются как композиции элементов и цифр с использованием встроенной Периодической системы элементов или по заранее введенным коэффициентам. Есть возможность регулировать величину поля – при XY 0-100% (Standard) и по величине поля, занятого данными (Auto).

18 – Диаграмма XYZ – оси выбираются или назначаются как композиции элементов и цифр . с использованием встроенной Периодической системы элементов или по заранее введенным коэффициентам. Есть возможность регулировать величину поля – при XYZ 0-100% (Standard) и по величине поля, занятого данными (Auto).

19 – Построение спайдер-диаграмм – относительных анализов. Выбирается эталонный анализ, на который делятся анализы файла.

20 – Перечень стандартных диаграмм, из которых можно выбирать «любимую». Возможно пополнение списка. (за счет построения пользователем произвольных диаграмм)

21 – Среднее, min, Мах – по данным помеченного файла (статистические характеристики выборки. дисперсия, асимметрия, эксцесс, гистограмма, аппроксимация Гауссом и т.д.)

22 – Статистические данные о материале файла

23 – Коэффициенты корреляции (КК) с указанием количества компонентов, учитываемых в коррелируемых парах элементов, и средних модулей КК для элементов

24 – Расстояния, расхождения ("расстояние" – общее название с "расхождениями", в которых не сохраняется неравенство треугольника) вызываются через кнопку 24. Есть два варианта: 1) по умолчанию – таблица расстояний «всех от всех» и 2) расстояния от выбранной точки в порядке анализов в файле. После 1) ОК в строке Меню появится «Distances», с перечнем 4-х вариантов расстояний: Минковского, евклидова, энтропийного, анэнтропийного.

25 – Стандартные петрологические коэффициенты

26 – Нормативные петрологические пересчеты.

27 – Классификация.

28 – Таблица ***RnEnAnTnRNEAT*** (сокращённо ***RHA***). Где: ***Rn*** – ранговая формула до ***n***-ного ранга; ***En*** – энтропия, стандартизованная при детальности ***n***; **An** – анэнтропия, стандартизованная при детальности ***n***; ***Tn*** – толерантность, стандартизованная при детальности ***n***; ***RN***– "хвост" ранговой формулы до***N***; ***E*** – энтропия полного анализа; ***A*** – анэнтропия полного анализа; ***T*** – толерантность полного анализа. Строки выдаются по умолчанию в «алфавитном порядке" – согласно алфавиту – Периодической системе элементов. При одинаковости ***Rn***упорядочивание производится по невозрастанию ***En***. При одинаковости ***En***упорядочивание производится по неубыванию ***An***, то же для ***Tn***. Для переупорядочения в исходный (авторский) порядок путь: Action-Order by source num+analysis num

29 – Выход т из программы.

**Идём по номерам ВЕРХНИХ заголовков раскрывающихся списков**

Некоторые заголовки **раскрывающихся списков** изменяются при конкретных действиях, т.е. за заголовком могут появляться другие заголовки.

1. Project Exit выход.

 После окончания работы по понедельникам спадает сообщение о создании очередной копии базы данных



1. View – виды представления данных – работает после выделения файла в 1 и 2 (нижние номера.
2. Results – перечень тех же кнопок
3. Data – дублирует ряд кнопок
4. Window – управление расположением материалов на рабочем столе
5. Option –управление работой программы. В открывшемся окне второе Options – крайняя правая вкладка ***RHA***. Здесь задание ***n*** детальности при расчетах – длины стандартной ранговой формулы, алфавиты, элементы, на которые делаются поправки в силикатных анализах при раздельном анализировании.
6. Double сlick add to list – Используется при составлении списков по кнопке 3 при добавлении нового анализа в список
7. About – Сведения о программе

**Поскольку пополнение базы данных может сопровождаться вводом данных об источнике информации, который фиксируется в библиографической Базе, освоение программы начнём с работы этого типа.**

**2. Ввод ссылки на источник данных**

Каждый файл должен иметь адрес возникновения – источника данных. Эта информация вводится в Библиографическую Базу

1. Вызов окна Библиографической базы (ББ): 4-я иконка меню слева 



По клику кнопки  откроется следующее окно



1. В открывшемся окне внизу находится шаблон записи. Вводим библиографические данные:

Number – Номер источника даётся в порядке поступления материалов в базу или как-то иначе.

Ф.И. О. автора(ов) или редактора («Ред.») (не более 80 символов)

Title – Название (не более 255 символов.):

Bibliogr. data – Издание – источник (не более 80 символов). Желательно указывать дату записи.

Если автор Вы, тогда: Ваша фамилия И.О. Название файла. ССА (серия собственных анализов). Полезно обозначение проекта, к которому относится файл.



1. При необходимости вызвать по номеру *уже внесенный источник* см. поле вверху – “Search by number” вводим номер. При необходимости редактируем текст.

Другие действия по поиску источника в Библиографической базе выполняются через «Запрос» = “Query”

**3. Требования к массиву анализов**

Качество подготовки данных – одно из основных условий успешной работы с программой. В связи с интегральностью получаемых характеризаций (количественных характеристик) составов, к исходным данным предъявляются повышенные требования, что является особенностью метода и, соответственно, подготовки материала для работы с программой.

***Требования к именам компонентов и использующиеся варианты***

Символ алфавита должен отвечать конкретному компоненту.

Поэтому компонентами должны быть или дискретными, или результатом дискретизации (верхние границы интервалов возрастов, длин волн, площадей...).

В качестве символов компонентов в ***химических анализах*** используются символы химических элементов (H, Si, Se...), оксидов (SiO2, H2O, Fe2O3...) и некоторых распространённых, простых по составу, соединений. Они приведены в таблицах – путь: Options-Options-Input Options.

Использование иных знаков, включая, ppm, %, LOI, TR, "+", "-" и прочие – запрещено.

Валентные состояния программа различает только для железа, концентрации для остальных элементов рассчитываются как суммарные. Чтобы при вводе данных программа правильно определяла валентность железа в химических соединениях, необходимо указать формулы этих соединений в списках на вкладке "Опции ввода" в диалоге настройки программы. Каждый компонент анализа может содержать железо только в одной из степеней окисления (исключением является Fe3O4, который программа интерпретирует правильно).

Алфавит ***химических элементов***, учтённых в программе (по умолчанию) – Периодическая Система Элементов приведён в разделе по пути: Option-Option-Alphabet №0

В ***минеральных составах*** используются аббревиатуры названий минералов с стандартизированной длиной – 4 (Dolm, Albt, Turm...), или, по желанию, пользователь сам вводит алфавит по Кретцу с отбором 100 важнейших минералов для изучаемой группы пород.

В ***кристаллохимических составах*** в качестве символов компонентов (в качестве примера алфавита введена символика для турмалина) используются сочетания символа позиции в структуре и элемента, встречающегося в ней. Для других минералов требуется формирование своих алфавитов ДО формирования общего для всех (возможен ли?).

Число компонентов в анализе не должно превышать 50. При этом речь идет именно о количестве компонентов в анализе. Число химических элементов в использующемся алфавите – 92, равное числу встречающихся в природе элементов.

При подготовке исходной таблицы для сканирования необходимо обращать внимание на точность обозначения одинаковых по начертанию букв в латинице и кириллице (С, О, К, Cа...), особенно – на сходство знака элемента кислорода "O" и цифры "0", а также на суммы анализов.

В обозначениях компонентов использование цифр недопустимо. Исключение – единственное – Масс% оксидов.

Совпадение символов в разных алфавитах допустимо.

***Требования к мерам содержаний***

Меры содержаний в пределах файла должны быть одинаковыми. В одном анализе недопустимо использование, например, весовых и молекулярных долей. В химических анализах допустимо совместное использование Mass% и ppm. ***Программа преобразует*** ***исходные данные*** весовых единиц оксидов в одинаковые, а именно ***в атомные содержания*** согласно условию: ∑***p***i=1. ***Именно такая форма*** представления химических составов ***обеспечивает универсальность метода*** по учёту любых компонентов, то есть отсутствие препятствий для сопоставления составов любых веществ, составов биообъектов, материалов и их смесей в одной таблице или на одной диаграмме. Более того. ***Составы, приведённые в такой*** ***форме, сопоставимы с любыми иными составами как статистических распределений нормированных к 1 или 100%..***

ВНИМАНИЕ! ***Требование к содержимому цифровых данных*** исходнойтаблицы

***Десятичный разделитель - точка.***

В ячейках таблицы допустимы только цифры.

***Требования к полноте анализов***

В анализе должны быть все компоненты, содержание которых превышают содержания наименьшего компонента из числа важнейших ***n***. В геологии особое внимание следует обращать на летучие компоненты. Как показано в «R- словаре-каталоге…»[[1]](#footnote-1) они входят в более чем половину всех известных минералов. Имея в виду и значимость "летучих" в вулканологии, в процессах метасоматоза и гидротермального переноса рудных компонентов, в процессах выветривания, традиционное игнорирование этих компонентов следует считать свидетельством затянувшегося отставания геологов в понимании значимости воды в геологии**.**

ИМЕТЬ В ВИДУ: Низкие суммы анализов (явление весьма обычное в минералогии и петрографии) ‒ проявление неполноты данных с возможными ***пропусками элементов, имеющих содержания большие, чем минимальные*** в имеющемся перечне анализа. В таких случаях расчёты интегральных характеризаций дают искажённые результаты.

При отсутствии необходимого алфавита, или при наличии компонентов с длиной символов, выходящих за пределы 9 знаков, при единичных расчётах можно пользоваться химическим алфавитом, заменяя его знаками необходимые при подготовке материалов к расчётам. После получения результата производится обратная замена знаков.

***Требования к описанию анализа объекта***

Описание анализа делается по стандарту, который обеспечивает возможности линейного упорядочения поступающих материалов, согласно номерам в Библиографической Базе Данных (иконка 4 - Bibliogrphy), а также поиски в Базе аналитических данных (Data sets) и Базе запросов Quories (путь: иконка меню №1 – Add – RHA). Пользователям, работающим со своими данными, или интересующимся не библиографией, а, например, географией, геологией и т.д. может быть полезными другие принципы описания и упорядочения.

В тексте "Описания" (Description или Descr) объекта обязательный порядок следующий: "номер анализа в таблице", например, «10»\_ номер библиографической ссылки, например, «960» значок "т" (сокращение слова *таблица*) или "р" (сокращение слова *page*) и номер таблицы или страницы\_"название объекта" (если есть) или "б/н" (если названия нет). Пример описания того же анализа: "10\_960-т21\_коматиит".

Далее подробности ‒ по потребности. Имя объекта с необходимыми свойствами приводится строчными буквами, заглавные используются только для имён собственных.

Предельная длина описания объекта 255 символов

Сокращения названий объектов, массивов и др. необходимо минимизировать (иначе при поиске аналогов по отдельным признакам будет высокий уровень шумов).

**4. Формирование файла**

Таблица формируется в Excel[[2]](#footnote-2).

Первая строка: Описание. "Description или Descr".Описания к массиву формируются в серии колонок как показано на скрине, после чего производится их сцепление в одну[[3]](#footnote-3), -



Далее в первой строке размещается перечень символов компонентов. Порядок компонентов произволен – он не зависит от принятого алфавита. Алфавит используется для контроля за правильностью обозначений вводимых компонентов и для алфавитного упорядочения массивов строк, получаемых при работе программы.



4) Установка стандартизации детальности ***n***

Перед вводом необходимо ***проверить установку*** стандартной ***детальности*** – длины учитываемой части ранговой формулы для расчётов характеризаций ***HАТ***, или произвести её. Путь: Option-Option-Program options- RHA-method – окошко слева внизу.

Далее ***таблицу необходимо выделить, не допуская захвата лишних строк и колонок.***

Таблица готова к вводу. Страницу можно свернуть (скрыть)

**5. Ввод файла**

Находим “Import”  (кнопка 6). Клик.

Имеем:



Для ввода готовой таблицы клик ОК, чем включатся *МАСТЕР ввода данных* как серия окон, фиксирующих последовательность шагов, обеспечивающих контроль за качеством исходных материалов и получение искомого результата.



Для продолжения : “Next”.

Имеем



ВНИМАНИЕ! При использовании нехимического алфавита нужно в окне Alphabet number поставить соответствующий номер. Для продолжения “Next”

 Получаем

:

Ошибки, обнаруженные в первой строке таблицы, выдаются в ЛЕВОМ поле (пустая колонка-строка… если пусто, ошибки на найдено). Обычны ошибки в символике компонентов. Для исправления вернуться – клик "Back" – в Excel. ***Перед возвращением в Petros не забыть выделить таблицу!***

Для продолжения “Next”

Имеем:



Если все поля пусты – ошибок на ЭТОМ этапе контроля не выявлено. Для продолжения “Next”.

Получаем:



Птичкой (по умолчанию) помечаются элементы, вводимые в виде ppm (миллионные доли), наряду с оксидами. Не допускать появления птичек перед символами элементов, содержание которых выражено в %%. Поскольку в примере введены только процентные величины, птички отсутствуют.

ВНИМАНИЕ! ***Если приведён заведомо не нормированный анализ,*** (например, весь в ppm) ***против всех символов элементов ставятся птички -*** (Tick all)

Для контроля цифровой информации клик “Validate”.

В СЛУЧАЕ погрешности в записи числа в спадающем окне появится сигнал Error.



В этом окне на фоне синей заливки в Error видна вторая десятичная точка. ЗДЕСЬ удалите её ‒ произойдёт ***её*** ***удаление и в исходных данных***. Проверка закончится.

Получаем сообщение в следующем виде:



Нажатием Next выходим на финишную прямую: **остаётся связать вводимый материал с Библиографической и с Аналитической базами Данных.**

Для продолжения “Next”.

Имеем: импорт данных из Excel.



В нижнем окошке «Number» по умолчанию появляется номер последнего в базе файла (здесь «2500»). Номер библиографической ссылки нового файла вводится в верхнее окно «Number». Под ним повторится номер и текст библиографии введённого первоисточника «1042». Этим организуется связь вводимого материала с введённой библиографической ссылкой.

В строке «Data set name» печатается номер источника с номером таблицы и краткое описание содержимого таблицы.

Если есть дополнения к описанию файла – соображения, акценты, комментарии, дата ввода… –текст вводится в следующую строку: Data set descripnion.

Производят общую проверку правильности введённой информации.

**ВНИМАНИЕ ! *Справа - не забывать проверить Data set tipe . Если исходные данные даны в Аt%, или коэффициентах химических формул, или их содержания измеряются количествами обнаружений*** – ***"штуками", то нажимается радиокнопка[[4]](#footnote-4)*** – ***Mol %.***

В таком виде Step 6 импорт произойдёт правильно



Клик “Finish”. Идет расчет.



Получаем исходные данные как свидетельство правильности их ввода:



ID – индивидуальный номер в БД, приписываемый анализу при вводе. Здесь первое число ‒ номер введённого файла, второй ‒ номер анализа в файле. Эти номера сохраняются при редактировании и удалении анализов **(!).**

Для фиксации результата в формирующейся базе данных Клик **ОК**.

**До «ОК» никаких иных** действий не производить!

Происходит выход на рабочее поле, свободное, если на нём не оставалось что-то от предыдущих действий.

**Поздравления! Материал введён!**

Он загружен в базу данных и готов к разностороннему использованию.

Для освоения ввода материала, рекомендуется эту процедуру сделать подряд несколько раз. Она представляется излишне сложной только на первый взгляд.

ПРИМЕЧАНИЕ:

Метод наилучшим образом проявляет свои положительные качества при наличии больших массивов данных, так как ЭТО:

1) делает более объективной оценку новизны и особенностей вашего материала, полноты, разнообразия, представительности, оригинальности

2) позволяет устанавливать сходство-различия материалов из разных источников данных, регионов, месторождений, различного генезиса…,

3) делает более осознанным выбор при формировании таблиц анализов для публикаций,

4) позволяет делать более обоснованные выводы,

5) расширяет кругозор и

6) будит воображение.

Если далее предстоит работа с только что введённым файлом, обратитесь к кнопке инструментальной панели Data set и в спустившемся окне вы обнаружите его выделенным. Для начала работы с файлом, уже имеющемся в базе, то есть в списке Data set, его нужно выделить. Далее, в обоих случаях, откройте окно Queries и в окне Data set двойным щелчком или нажатием кнопки справа внизу  , перешлите файл в окно запросов – Queries. Там появится выделенный вами файл, готовый к работе.

В качестве первого действия рекомендуется ознакомиться с главной таблицей ***RHAT***, для чего нажмите предпоследнюю (28-ю) кнопку инструментальной панели (по разнообразию ранговых формул можно оценить степень однородности материала, представленность разных групп, выявить оригинальные и банальные составы). Также полезно для общего представления об особенностях материала вызвать диаграмму EnAn (на ней представлена форма поля-полей составов – его или их изо-анизометричность, равномерность распределения точек в поле, отскакивающие точки…). Диаграмма ***EnAn*** заранее выбирается как «Любимая» из группы появляющихся при нажатии кнопки 20.

Подавляющее большинство действий программы выполняются только при наличии файла в списке Queries и его выделения там заливкой или птичкой.

Файл после окончания работы остаётся в списке под кнопкой Queries и может впоследствии вызываться без обращения к Data set.

**Успешного освоения программы и получения интересных результатов!**

1. Петров Т.Г., Краснова Н.И. R-словарь-каталог химических составов минералов. СПб, «Наука», 2010, 150 с. Грант РФФИ № 09 – 05 – 07070д [↑](#footnote-ref-1)
2. Приводится описание авторского опыта и способа работы с программой. [↑](#footnote-ref-2)
3. ***Сцепка частей описания в разных колонках в одну*** производится в Excel по кнопке "Формулы" - "Вставить функцию"-"Сцепить" ОК. Открытие окна "Аргументы функции", выделение первой ячейки в первой колонке элементов описания, перенос курсора на вторую строчку- выделение первой ячейки во второй колонке элементов описания, так далее... ОК. В Excel протащить первую строку до конца колонки. "Копировать"- вызов: "специальная вставка". В открывшемся окне клик "Значения", ОК. [↑](#footnote-ref-3)
4. Радиокнопка нажата:  [↑](#footnote-ref-4)