

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
ИНСТИТУТ ФИЗИКИ ЗЕМЛИ ИМ. О.Ю. ШМИДТА
РОССИЙСКОЙ АКАДЕМИИ НАУК

на правах рукописи



Иванов Станислав Дмитриевич

**Применение информационных технологий
в геобаротермометрии и смежных задачах**

Специальность:

25.00.10 — «Геофизика, геофизические методы поисков полезных
ископаемых»

Диссертация на соискание ученой степени кандидата технических наук

Научный руководитель
кандидат физико-математических наук
И. М. Алешин

Москва — 2017

Оглавление

Используемые сокращения	5
Введение	6
1. Компьютерные системы для обработки петролого-геохимических данных	12
1.1. Петролого-геохимические данные и методы их обработки	12
1.2. Управление геохимическими данными и их качественный анализ	13
1.3. Идентификация минералов	17
1.4. Определение слагающих породу минералов по результатам валового анализа	18
1.5. Геобаротермометрия	20
1.6. Базы геохимических данных	23
1.7. Выводы	25
2. Интерактивный реестр методов обработки петролого-геохимических данных на основе технологии вики	27
2.1. Концепция интерактивного реестра	27
2.2. Интерактивный реестр на основе веб-приложения	29
2.2.1. Технология вики	32
2.2.2. Выбор вики-платформы	33
2.2.3. Организация информации о методах в рамках вики	35
2.2.4. Организация работы с библиографией в рамках вики	38
2.3. Методы обработки геохимических данных в интерактивном реестре	40
2.3.1. Управление пользовательскими данными	43
2.3.2. Выбор языка программирования для описания методов	45

2.3.3.	Программный интерфейс методов	47
2.3.4.	Правила представления данных для их однозначной интерпретации	48
2.3.5.	Управление методами	49
2.3.6.	Обеспечение безопасности при выполнении метода .	50
2.3.7.	Пользовательский интерфейс выполнения расчетов . .	51
2.3.8.	Поддержка многоязычности	53
2.4.	Инструменты визуализации геохимических данных в интерактивном реестре	54
2.4.1.	Базовые инструменты визуализации	54
2.4.2.	Специализированные инструменты визуализации петролого-геохимических данных	57
2.4.3.	Интеграция инструментов визуализации с интерактивным реестром	59
2.5.	Работа с методами вне реестра	61
2.6.	Выводы	61
3.	Интерактивный реестр методов геобаротермометрии	64
3.1.	Формальное описание геосенсора	64
3.2.	Реализация на базе интерактивного реестра	65
3.3.	Наполнение базы методами	70
3.3.1.	Импорт базы сенсоров программы TRF	70
3.4.	Использование реестра на примере фенгитовой мономинеральной барометрии	72
3.5.	Выводы	76
4.	Система идентификации минералов по результатам химического анализа	77
4.1.	Идентификация минералов средствами нечеткой логики . . .	77
4.1.1.	Определение параметров	81
4.2.	Программная реализация	82
4.3.	Автоматическое составление наборов ограничений	83
4.4.	Интеграция с интерактивным реестром	84
4.5.	Использование системы	86
4.6.	Выводы	87

Заключение	89
Литература	91
Список иллюстраций	102
Список таблиц	103
Список листингов	104
Приложение 1. Основные операции в интерфейсе интерактивного	
реестра	105
Вход в систему	105
Главная страница реестра	106
Управление файлами пользователя	107
Корневая страница пространства имен	109
Базовая визуализация	110
Модуль специальной визуализации	111
Модуль геобаротермометрии	112
Модуль идентификации минералов	113

Используемые сокращения

<i>ОС</i>	— Операционная система
<i>СПО</i>	— Свободное программное обеспечение
<i>к.ф.</i>	— Коэффициенты формул
<i>ACL</i>	— англ. Access Control Lists, Списки контроля доступа
<i>AJAX</i>	— англ. Asynchronous Javascript and XML, Асинхронный JavaScript и XML
<i>API</i>	— англ. Application Programming Interface, Интерфейс программирования приложений
<i>CMS</i>	— англ. Content Management System, Система управления содержанием
<i>DOI</i>	— англ. Digital Object Identifier, Цифровой идентификатор объекта
<i>GNU GPL</i>	— англ. GNU General Public License, Общая публичная лицензия GNU
<i>HTTP</i>	— англ. HyperText Transfer Protocol, Протокол передачи гипертекста
<i>JSON</i>	— англ. JavaScript Object Notation, Нотация объектов в JavaScript
<i>MVC</i>	— англ. Model — View — Controller, Модель — Представление — Контроллер
<i>REST</i>	— англ. Representational State Transfer, Передача представления о состоянии
<i>TDD</i>	— англ. Test-driven Development, Разработка через тестирование
<i>URL</i>	— англ. Uniform Resource Locator, Единый указатель ресурса
<i>VCS</i>	— англ. Version Control System, Система управления версиями
<i>ppm</i>	— англ. Parts per million, Миллионная доля

Введение

Петролого-геохимические данные, в частности, валовые составы горных пород и составы слагающих их минералов являются важным источником данных для исследования эволюции горных пород [Spearg, 1995]. На сегодняшний день предложено множество методов их обработки и интерпретации [Скляр, 2001]. Среди них можно выделить методы предварительной обработкой данных (геохимические и минералогические пересчеты, расчет характеристических коэффициентов), их качественной обработки (дискриминационные диаграммы и т.п.) и методы получения конечных численных оценок, таких как давление и температура, на различных этапах эволюции горной породы (геотермометры и геобарометры). Эти методы широко применяются в геологических исследованиях [Сизых, Буланов, 1991].

Методы обработки петролого-геохимических данных являются хорошо алгоритмируемыми, что обеспечивает широкое применение информационных технологий для их обработки. Программное обеспечение, реализующее эти методы, представлено широким спектром решений от интегрированных программных пакетов до электронных таблиц. Можно выделить несколько классов таких программ. Это настольные приложения и электронные таблицы для расчетов, связанных с валовым составом горных пород ([Janousek, 2006], [Carr, 2002], [Su, Langmuir, Asimow, 2003] и др.), минералогическими пересчетами ([Brey, Bulatov, Girnig, 2008], [Preston, 2011], [Horn et al., 2013], и др.), идентификацией минералов ([Perroud, 2013], [Smith, Leibovitz, 1986] и др.) и решения задач геобаротермометрии ([Fonarev, Graphchikov, Konilov, 1991], [Кориневский, 2015], [Sturm, 2002], [Yavuz, 2013] и др.).

Несмотря на разнообразие существующих программ для обработки петролого-геохимических данных, можно выделить некоторые проблемы, присущие в большей или меньшей степени всем им. В первую очередь, это недостаток справочной информации о методах, которая предоставляется пользователю. Справочная информация о них в приложениях, реализующих многие десятки методов, например PET [Dachs, 1998] или TPF [Fonarev, Graphchikov, Konilov, 1991], ограничена литературной ссылкой, которой

недостаточно для обоснованного выбора пользователем того или иного метода [Иванов, 2015b].

Другой проблемой является закрытость исходных кодов методов и зависимость большинства программ от платформы, что затрудняет воспроизведение результатов и ограничивает возможность пользователей делиться промежуточными результатами. Среди открытых программ — коллекций методов обработки петролого-геохимических данных следует отметить GCDKit [Janousek, 2006], однако и этот продукт является платформозависимым.

Еще одна проблема состоит в невозможности пополнения коллекций методов пользователями. Попытка ее решения была предпринята в программе TPF [Fonarev, Graphchikov, Konilov, 1991], однако широкого распространения этот подход не получил. С появлением новых методов и выходом публикаций об уже существующих, встает вопрос об обновлении программ для актуализации информации.

Кроме того, при решении прикладных задач обработки петролого-геохимических данных встает проблема определения минерала по его химическому составу, которая на сегодняшний день недостаточно разработана [Иванов, 2016с].

Таким образом, **актуальным** является создание компьютерного инструмента, который позволил бы не только организовать и использовать существующее множество методов обработки петролого-геохимических данных, но и помог бы специалисту-геологу с выбором конкретного метода.

Целью работы является создание программного пакета, объединяющего в себе справочную информацию о методах обработки петролого-геохимических данных, их алгоритмы и инструменты выполнения расчетов. Этот пакет получил название интерактивного реестра. Термин «реестр» вводится в данной работе, чтобы подчеркнуть отличия данного пакета как от программных библиотек методов, так и от традиционных текстовых справочников. Кроме того, этот термин подразумевает возможность дополнения и развития как самих методов, так и информации о них. «Интерактивность» реестра подразумевает возможность непосредственного использования имеющихся в нем методов через встроенный в приложение пользовательский интерфейс.

Для достижения этой цели были поставлены следующие **задачи**:

- Анализ существующих решений в области обработки петролого-геохимических данных;
- Выбор архитектуры и платформы для реализации программного пакета, его разработка;
- Реализация на его основе вспомогательных инструментов обработки петролого-геохимических данных;
- Разработка системы идентификации минералов и ее интеграция с приложением;
- Реализация инструмента выполнения расчетов в области геобаротермометрии;
- Использование созданного пакета для решения прикладных задач.

Научная новизна данной работы состоит в использовании нового подхода к организации методов обработки петролого-геохимических данных и справочной информации о них в рамках веб-приложения, получившего название интерактивного реестра. Этот подход позволяет повысить информационную обеспеченность петролого-геохимических исследований, обеспечить прозрачность реализации методов [Prlić, Procter, 2012], а также использовать преимущества коллективной разработки [Andreadisa, Fourtounisb, Bouzakisa, 2015].

Теоретическая значимость данной работы состоит в применении нового подхода к организации методов расчета, в рамках которого не происходит разделение описания и алгоритма метода. Кроме того, показана возможность использования методов нечеткой логики для решения задачи идентификации минералов по химическому составу. **Практическая значимость** работы состоит в создании инструмента — научного веб-приложения для решения задач обработки петролого-геохимических данных. В частности, использование созданного инструмента позволило получить ценные научные результаты в работе по исследованию блыбского метаморфического комплекса Передового хребта Северного Кавказа [Камзолкин, Иванов, Конилов, 2015].

В рамках поставленных задач были изучены существующие компьютерные системы для обработки петролого-геохимических данных, работающие с валовыми составами и составами отдельных минералов, инструменты

геобаротермометрии, системы идентификации минералов, базы геохимических данных. Сформулирована концепция интерактивного реестра, выбрана его архитектура и предложено использовать технологию вики для организации данных о методах. Выполнено сравнение современных вики-платформ с точки зрения использования в качестве основы для построения реестра. Итогом этого сравнения стал выбор платформы DokuWiki [Victorino, 2012]. Создано веб-приложение — интерактивный реестр методов обработки петролого-геохимических данных. В качестве модулей к нему были разработаны и реализованы инструменты для выполнения петролого-геохимических расчетов и управления данными. Осуществлено наполнение реестра методами, представленными в литературных источниках.

По итогам разработки, реализации и апробации интерактивного реестра **на защиту выносятся следующие положения:**

1. Концепция интерактивного реестра методов обработки петролого-геохимических данных позволяет организовывать и систематизировать разнородную информацию о самих данных, а также о методах и алгоритмах их обработки в рамках единой иерархической информационной системы.
2. Веб-приложение, которое объединяет в себе справочный материал, алгоритмы методов и инструменты для выполнения расчетов в геобаротермометрии и смежных задачах, реализует предложенную концепцию интерактивного реестра.
3. Проблемно-ориентированная система идентификации минералов на основе нечеткой логики, реализованная на базе интерактивного реестра, которая решает задачу определения минералов по результатам химического анализа в рамках имеющегося набора правил.

Апробация результатов диссертационной работы. Основные результаты были опубликованы в 11 работах. В том числе в 3 статьях в журналах, включенных в перечень российских рецензируемых научных журналов и издания для опубликования основных научных результатов диссертации:

1. *Иванов С.Д.* Интерактивный реестр геосенсоров на основе веб-приложения // Компьютерные исследования и моделирование. — 2016а. — Т. 8, № 4. — С. 621–632.

2. *Иванов С.Д.* Метод идентификации минералов по химическому составу на основе нечеткой логики // Геофизические исследования. — 2016с. — Т. 17, № 4. — С. 47–56. — DOI: 10.21455/gr2016.4-4.
3. *Камзолкин В.А., Иванов С.Д., Конилов А.Н.* Эмпирический фенгитовый геобарометр: Обоснование, калибровка и применение // Записки Российского Минералогического Общества. — 2015. — Т. 144, № 5. — С. 1–14.

На приложение для идентификации минералов по химическому составу на основе нечеткой логики получено авторское свидетельство:

4. *Свидетельство о государственной регистрации программы для ЭВМ № 2016616340 Российская Федерация.* Программа для идентификации минералов по химическому составу на основе нечеткой логики / *С.Д. Иванов* ; заявитель и правообладатель ИФЗ РАН. — № 2016613557 ; заявл. 13.04.2016 ; опубл. 09.05.2016.

Результаты диссертационной были отражены в тезисах докладов и материалах конференций:

5. *Иванов С.Д.* Современная платформа для обработки результатов геохимических анализов // Физико-химические и петрофизические исследования в науках о Земле. Материалы XV международной конференции. — М.: ГЕОХИ, 2014а. — С. 92–94.
6. *Иванов С.Д.* Экспертная система определения минералов по результатам микронзондовых анализов // Конференция молодых ученых ИФЗ РАН. — Москва: ИФЗ РАН, 2014б. — С. 44.
7. *Иванов С.Д.* Интерактивная база методов петрохимической обработки данных на основе технологии вики на примере системы идентификации минералов // Конференция молодых ученых ИФЗ РАН. — Москва: ИФЗ РАН, 2015а.
8. *Иванов С.Д.* Некоторые требования к интерактивной базе методов обработки геохимических данных // Науки о Земле. Современное состояние. Материалы третьей всероссийской молодежной научно-практической школы-конференции. — Шира: НГУ, 2015б. — С. 119–120.

9. *Иванов С.Д.* Интерактивный реестр методов обработки петролого-геохимических данных // Научная конференция молодых ученых и аспирантов ИФЗ РАН. Тезисы докладов и программа Конференции. — Москва: ИФЗ РАН, 2016b. — С. 33.
10. *Камзолкин В.А., Иванов С.Д., Конилов А.Н.* Некоторые результаты анализа зависимости состава белых слюд (фенгита) от P-T-условий их формирования // Физико-химические и петрофизические исследования в науках о Земле. Материалы XIV международной конференции. — М.: ГЕОХИ, 2013. — С. 123–126.
11. *Ivanov S.D. et al.* Chemical and mineralogical data and processing methods management system prototype with application to study of the North Caucasus Blybsky Metamorphic Complexes metamorphism P-T-condition // EGU General Assembly 2014, Geophysical Research Abstracts. Vol. 16. — Viene: EGU, 2014. — P. 845.

Разработанная система была использована в рамках исследования блыбского метаморфического комплекса Передового хребта Большого Кавказа [Камзолкин, Иванов, Конилов, 2015].

Система была развернута на виртуальной серверной инфраструктуре ИФЗ РАН и доступна в сети Интернет по адресу <http://pin.ifz.ru>.

Глава 1. Компьютерные системы для обработки петролого-геохимических данных

1.1. Петролого-геохимические данные и методы их обработки

Основу петролого-геохимических данных составляют сведения о составе горных пород и слагающих их минералов. Состав включает относительное или абсолютное содержание отдельных элементов, содержание элементов в различных степенях окисления и содержание и соотношение отдельных изотопов. Эти данные могут дополняться информацией об образце, горной породе, минерале, точке, аналитической сессии и настройках аналитического оборудования, а также условиях эксперимента для экспериментальных образцов. Дополнительная информация носит разнородный и слабо формализуемый характер, в отличие от данных о химическом составе, которые выражены, как правило, в весовых процентах оксидов, атомных весовых процентах (ppm, ppb), атомных процентах или коэффициентах формул. Петролого-геохимические данные можно разделить на валовые, описывающие состав породы в целом, и детальные, описывающие состав отдельных минералов, слагающих породу. Они имеют типично табличную структуру, в качестве основного ключа используется номер образца или же пара номер образца – номер анализа. Результаты анализов объединяются в наборы по различному (и, возможно, произвольному) признаку. Например, пробы из одного образца, пробы одного минерала, пробы одной экспедиции, пробы одной сессии микрозонда и т.д. (подробнее о метаданных см. раздел 1.6).

Методы обработки петролого-геохимических данных можно условно разделить на качественные и количественные. Качественные методы, как правило, подразумевают получение некоторого визуального представления данных, удобного для дальнейшей интерпретации. К ним можно отнести построение различных дискриминационных диаграмм и спайдерграмм. Количественные методы подразумевают получение конкретных численных оценок. В частности, к ним относятся методы геобаротермометрии и геохронологии. Эти методы являются хорошо алгоритмируемыми

и используют однотипные исходные данные. Однако их разнообразие чрезвычайно велико, а применимость ограничена трудноформализуемыми, а подчас и неформализуемыми, ограничениями. Кроме того, в качестве отдельной задачи следует отметить управление самими данными.

Отдельно необходимо выделить методы, лежащие в основе широко представленных в литературе систем термодинамического моделирования. Среди них выделяется класс программных продуктов для работы с геологическими системами, к которому относятся такие программы, как *TWQ* [Berman, 2007], *Thermocalc* [Powell, Holland, Worley, 1998], *Perplex* [Connolly, 2009], *Theriak-domino* [Capitani, Petrakakis, 2010], библиотека *arxim* [Moutte, 2009] и др. Методы, лежащие в их основе, являются существенно более сложными, чем те, что упомянуты выше. Прежде всего, это методы, связанные с расчетом баланса фаз, их числа, макропараметров, условий термодинамических переходов и т.д. Системы термодинамического моделирования (за редкими исключениями, например, *PTGibbs* [Brandelik, Massonne, 2004]) представлены в виде отдельных программных пакетов или библиотек, неразрывно связанных с базами термодинамических моделей веществ и твердых растворов, на основе которых и выполняется моделирование. Рассмотрение их специфики выходит за рамки данной работы.

Программные продукты, реализующие рассматриваемые методы обработки петролого-геохимических данных, представлены в виде электронных таблиц, наборов скриптов для настольных математических пакетов, а также отдельных приложений. Существующие программные продукты можно разбить на несколько категорий в зависимости от решаемых ими задач. Следует отметить условность такого деления, так как множества задач, решаемых различными программными пакетами, существенно перекрываются, поэтому представленное деление вводится для структурирования повествования, а не на правах классификации.

1.2. Управление геохимическими данными и их качественный анализ

Желание использовать компьютер для работы с петролого-геохимическими данными закономерно, учитывая их числовую природу. Поэтому в первую очередь следует отметить программы, претендующие на роль

настольного инструмента ученого-петролога. Эти приложения объединяют в себе графический пользовательский интерфейс для управления данными и определенный набор методов их обработки. В архитектуре таких приложений можно выделить два основных подхода: самостоятельные приложения и расширения для различных настольных пакетов.

Одним из первых самостоятельных приложений с графическим пользовательским интерфейсом этого класса является программа *MinPet* [Richard, 1995]. Более поздними разработками стала программы *PetroGraph* [Petrelli et al., 2005] и отечественная разработка *PetroExplorer* [Кориневский, 2015] (рис. 1.1). Основные задачи, для решения которых созданы данные приложения, это построение дискриминационных диаграмм и спайдерграмм, произвольных диаграмм, тригонограмм и выполнение геохимических расчетов. Кроме того, расчет CIPW нормы и решение задачи баланса массы (см. раздел 1.4) в *PetroGraph* и минералогические пересчеты и использование небольшого набора геотермометров в программе *PetroExplorer*. Эти продукты являются приложениями для операционной системы Microsoft Windows и распространяются бесплатно, однако их исходные коды являются закрытыми. Для хранения геохимических данных *PetroExplorer* использует для этого базу данных Microsoft Access, в то время, как *PetroGraph* сохраняет данные в собственном формате и импортирует геохимических данных из электронных таблиц средствами Microsoft Excel. Следует упомянуть также коммерческую программу *IgPET* [Carr, 2002], не вошедшую в обзор, т.к. она не доступна для свободной загрузки. Развитие всех упомянутых выше проектов на данный момент не ведется, однако их финальные версии по-прежнему представляют ценность в качестве относительно удобных инструментов для обработки петролого-геохимических данных. К сожалению, остановка развития проекта при отсутствии открытых исходных кодов рано или поздно приведет к невозможности их использования в современных операционных системах.

Второй ветвью настольных петрологических систем являются приложения, построенные на базе настольных пакетов. Для создания геохимических диаграмм в пакете Microsoft Excel существует надстройка под названием *PetroPlot* [Su, Langmuir, Asimow, 2003]. Для настольного математического пакета Wolfram Research Mathematica [Haneberg, 2012] был разработан пакет *PET* [Dachs, 1998]. А на основе пакета MathWorks Matlab [Hunt, Lipsman,

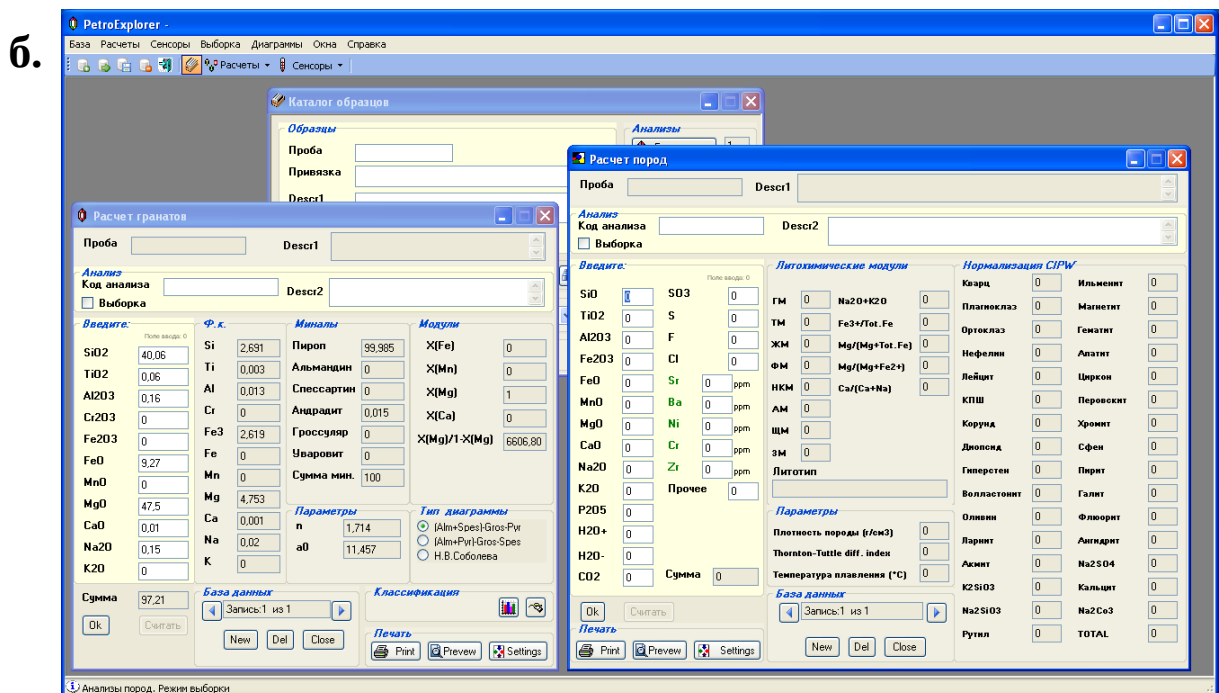
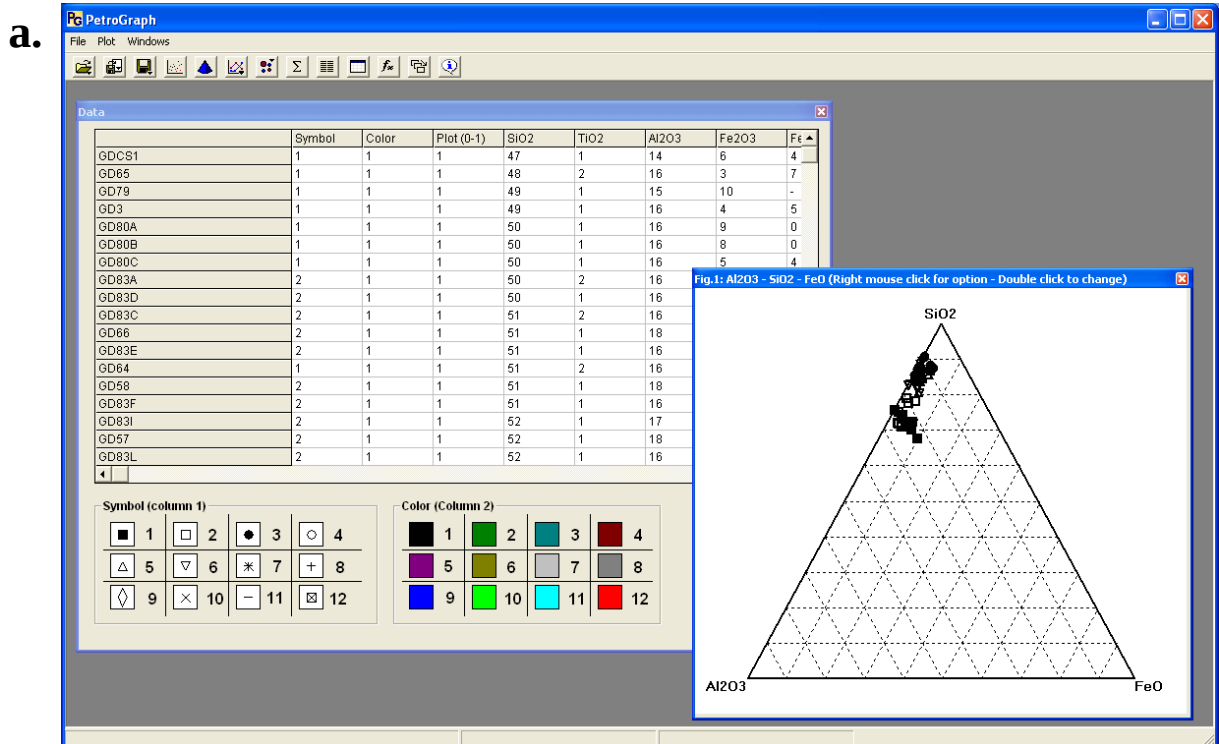


Рис. 1.1. Интерфейс приложений *PetroGraph* [Petrelli et al., 2005] (а) и *PetroExplorer* [Кориневский, 2015] (б). Показаны главные окна приложений, меню, панель инструментов и некоторые доступные диалоговые окна

Rosenberg, 2014] был создан проект *FastGAPP* [Riefstahl, 2012]. Стремление авторов программного обеспечения для обработки петролого-геохимических данных использовать существующие математические платформы закономерно — такой подход позволяет избежать реализации базовых примитивов по работе с данными (представление данных пользователю в виде таблицы, поиск, загрузка и сохранение файлов и т.п.) и построения графиков (современные пакеты предоставляют обширный набор готовых графиков и диаграмм), т.к. они доступны в рамках платформы в виде готового API. Обратной стороной данного подхода является зависимость программы от используемого пакета (а в случае *PetroPlot* — от конкретной версии Microsoft Office), что в случае, если он не является свободным ПО, может ограничить число пользователей.

Особое внимание хочется уделить проекту *GCDKit* [Janousek, 2006] (рис. 1.2), разработанному и поддерживаемому Янусеком Войтечем. Данный проект представляет набор модулей на языке R, ориентированный на работу с петрохимическими данными по изверженным породам. *GCDKit* интегрируется с интерфейсом RConsole (только в операционной системе Microsoft Windows) добавляя возможность вызова основных инструментов в меню. Данный продукт позволяет строить несколько десятков геохимических диаграмм, спайдерграммы и выполнять некоторые геохимические пересчеты. *GCDKit* поддерживает несколько десятков классификационных диаграмм горных пород (AFM, TAS, и др.) и определения тектонических обстановок. Этот пакет использует возможности R для экспорта и импорта, что обеспечивает возможность загрузки геохимических данных из большинства популярных форматов (XLS, CSV, базы данных Microsoft Access и др.).

Пользователь *GCDKit* может использовать как интерактивную консоль для вызовы функций, так и основанный на меню и диалогах графический пользовательский интерфейс. В отличии от предыдущих продуктов *GCDKit* является свободным программным обеспечением и успешно развивается на сегодняшний день. Его исходные коды доступны в виде библиотеки на сайте проекта: <http://www.gcdkit.org/download>.

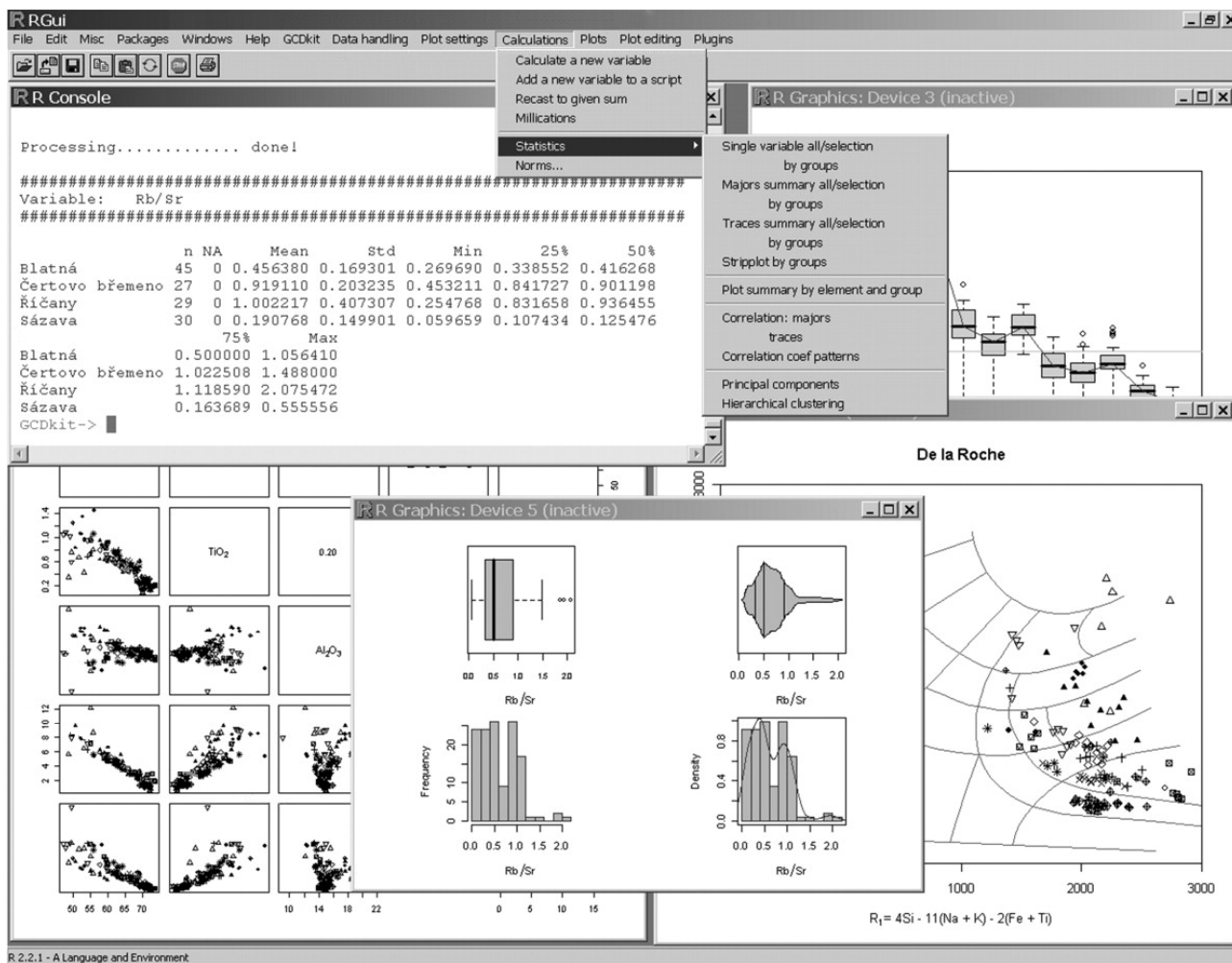


Рис. 1.2. Интерфейс GCDKit (из статьи [Janousek, 2006]). Показаны результаты работы некоторых доступных в программе инструментов визуализации и дополненное меню оболочки RConsole

1.3. Идентификация минералов

К информации о минеральном составе образцов предъявляются высокие требования в задачах геобаротермометрии. Кроме того, некоторые геосенсоры накладывают дополнительные ограничения на химический состав минералов. Поэтому возможность идентификации минералов и контроля минерального состава образцов после проведения микрозондового анализа является полезной.

Компьютерные системы для идентификации минералов представлены в первую очередь экспертными системами, предназначенными для помощи при ручном определении минералов в полевых условиях на основе общих физических и химических свойств (например, [Starek et al., 2002], [Folorunso et al., 2012] и др.). Эти системы не работают непосредственно

с геохимическими данными, а потому их рассмотрение выходит за рамки данной работы, также как и интерактивные ключи для идентификации, например [Plante, Peck, Barger, 2003]. В системе *ATHENA* ([Perroud, 2013]) реализована полуавтоматическая идентификация минералов с использованием интерактивного минералогического справочника. Данная система предоставляет возможность поиска минералов по списку содержащихся в них элементов и также не работает с численными данными.

В контекст настоящей работы входят те системы, которые используют количественные оценки химического состава минералов для определения их наименований. В литературе описаны различные подходы к решению этой задачи. Наиболее популярным является непосредственное сравнение результатов анализа с эталонным составом. Данный подход реализован в таких программах, как *MinIdent* [Smith, Leibovitz, 1986], *GemIdent* [Omoumi, 1990], *MINCALC* [Bernhardt, 2010]. Эталонный состав может быть рассчитан по теоретической формуле минерала или задан разработчиком. Еще один подход предложен в системе иерархической классификации РНА [Петров, 2001]. В ней для определения наименования исследуемого объекта используется только порядок элементов при сортировке их содержаний по убыванию.

Для существующих систем определение минералов с развитым изоморфизмом представляет сложность. Описание таких минералов через эталонный состав приводит к тому, что успешно могут быть определены только конечные члены рядов твердых растворов, что существенно снижает полезность подобных систем. Кроме того, эти системы не позволяют пользователям создавать и использовать собственные базы минералов. Эта возможность была бы полезна, например, при отборе данных для дальнейшего использования в геобаротермометрии.

1.4. Определение слагающих породу минералов по результатам валового анализа

Задача определения минерального состава пород по результатам валового химического анализа интересовала минералогов и геохимиков на протяжении многих десятилетий. Применение компьютеров позволило выйти в этой области на качественно новый уровень [Putirka, 2008].

К решению этой задачи существует два основных подхода. Первый подход состоит в решении «обратной задачи» для валового анализа. В рамках этого подхода компьютерная программа, исходя из известного состава породы и списка минералов, входящих в нее, пытается рассчитать состав отдельных минеральных фаз и их количественное соотношение. Такие программы, как *SEDNORM* [Cohen, Ward, 1991] и *LPNORM* [Caritat, Bloch, Hutcheon, 1994], решают задачу оптимизации баланса массы, находя лучшее решение для системы линейных уравнений в некотором пространстве ограничений. Программа *MODAN* [Paktunc, 2001] использует метод Гаусса, а потому требует, чтобы число минералов было меньше или равно числу компонентов состава, при этом решение в ней находится множественной регрессией на основе метода наименьших квадратов. Отечественная разработка *МС* [Чудненко, Авченко, Вах, 2013] также использует для решения метод наименьших квадратов и ориентирована на метасоматические породы. В отличие от перечисленных выше программ, *A2M* [Posch, Kurz, 2007] ищет не единственное решение, а вычисляет ортогональный базис пространства решений, что позволяет получить конкретное решение в виде линейной комбинации компонентов базиса. Очевидно, что с формальной точки зрения решение такого рода задач может быть не единственным, однако, привлекая дополнительные геологические сведения, возможен выбор наиболее вероятного минерального состава породы. Следует отметить, что с появлением современных методов микроанализа потребность в таких программах существенно снизилась ввиду возможности непосредственного измерения состава отдельных минеральных зерен горной породы.

Второй подход основан на так называемых нормативных пересчетах. В рамках такого подхода производится расчет минерального состава с использованием формального алгоритма для определенного класса горных пород. Классическим, но не утратившим до сих пор актуальность алгоритмом является вычисление *CIPW* [Kelsey, 1965] нормы, а также ее вариаций (например, [Дубровский, 2000] или алгоритма программы *Minlith* [Rosen, Abbyasov, Tipper, 2004]). Методы ее вычисления *CIPW* нормы входят во многие петрологические программы, рассмотренные в разделе 1.2. Следует отметить, что нормативные пересчеты выполняются

для некоторой идеализированной модели горной породы, поэтому полученный минеральный состав не всегда соответствует составу образца. Однако они остаются полезными инструментами при исследовании эволюции магматических горных пород.

1.5. Геобаротермометрия

Методы количественной оценки условий формирования горной породы на основе анализа химического состава минералов, слагающих ее, носят название геотермометры и геобарометры. Эти методы также называют геосенсорами [Симакин, Закревская, Салова, 2012], геобаротермометрами [Авченко и др., 2014]. Существует большое разнообразие геосенсоров для различных пород и минеральных ассоциаций. Их число на сегодняшний день исчисляется многими десятками. Несмотря на фундаментальность принципов, лежащих в основе минеральной геобаротермометрии, существует большая опасность некорректного применения ее методов. Прежде всего, любой сенсор предполагает равновесность той минеральной ассоциации, к которой он применяется [Аранович, 1991]. Также принципиальное значение имеет история химических реакций и процессов кристаллизации, происходивших в породе [Spear, 1995]. Кроме того, калибровки сенсоров, как правило, выполняются для конкретных горных пород, минеральных ассоциаций и диапазонов содержания химических элементов. Авторы сенсора, в свою очередь, могут накладывать дополнительные ограничения на его применение, в связи с чем может потребоваться дополнительное знакомство с первоисточником [Иванов, 2016а]. Следует отметить, что для одной минеральной ассоциации может существовать несколько различных геосенсоров, а для каждого сенсора может быть опубликовано несколько калибровок. Например, гранат-клинопироксеновый геотермометр, применяемый для оценки температур в эклогитах, высокобарных гранулитах, гранат-амфиболитах и гранат-перидотитах, предложенный Эллисом и Грином в 1979 году [Ellis, Green, 1979], совершенствуется до сих пор и насчитывает уже не менее двенадцати калибровок [Nakamura, 2009].

Программы, выполняющие расчет для отдельного геосенсора, представлены в основном в виде электронных таблиц, прилагаемых к статьям, в которых сенсор предлагается или рассматривается (например [MacGregor,

2015a)]. Для относительно небольшого числа геосенсоров существуют отдельные приложения, выполняющие соответствующий расчет.

В то же время, организованные тем или иным образом коллекции сенсоров создавались неоднократно в виде самостоятельных приложений (например, [Fonarev, Graphchikov, Konilov, 1991] и др.), отдельных инструментов в рамках приложений (например, [Кориневский, 2015]) и в виде электронных таблиц (например, [Brey, Bulatov, Girnis, 2008], [Preston, 2011], [Hora et al., 2013], [Nakamura, 2015] и др.). Относительная простота реализации многих методов геобаротермометрии в виде электронных таблиц обусловила широкое распространение их именно в этой форме. Электронными таблицами, использующими для своей работы механизм макросов, являются *Winpyrox* [Yavuz, 2013], *GPT* [Reche, Martinez, 1996] или *BGT* [Rameshwar, 1995]. Также следует выделить коллекции электронных таблиц, собранные различными авторами, например [Tindle, 2015] или [Waters, 2004]. В веб-приложении для термодинамических расчетов *Computational Thermodynamics Server* [Ghiorso, Kress, 2011] существует раздел, посвященный геосенсорам, в котором имеется веб-интерфейс к некоторым методам геотермобарометрии.

Среди имеющихся коллекций сенсоров следует отметить отечественную разработку — программу *TPF* [Fonarev, Graphchikov, Konilov, 1991]. Кроме большого набора доступных сенсоров (более 200), ее отличает архитектура? в рамках которой сенсоры не являются частью ее исходного кода, а описываются на специальном Pascal-подобном языке. Этот подход позволяет пользователю расширять базу программы новыми методами расчета. Кроме того, программа содержит ряд инструментов для работы с геохимическими данными (рис. 1.3). Следует отметить, что петрологические пересчеты в рамках данной программы являются жестко заданными, что не позволяет дополнять базу сенсорами, использующими специфические методы пересчета. Последняя версия программы была выпущена для операционной системы DOS в 2004 году. Она является 16-битной и может быть запущена на современных компьютерах только с использованием эмулятора (например, DOSBox).

Серьезной проблемой, стоящей перед специалистом-геологом, является выбор конкретного сенсора или набора сенсоров, который допустимо применять к его исходным данным. Несмотря на фундаментальность

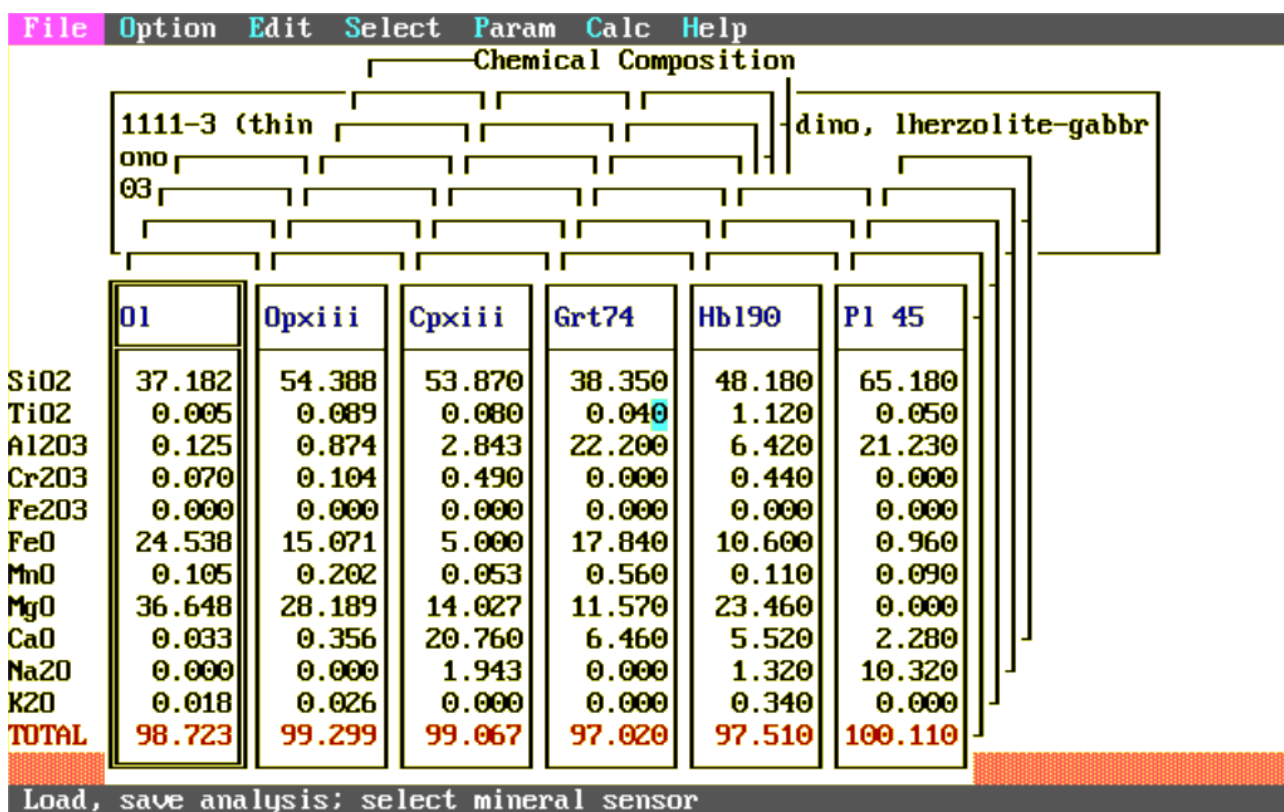


Рис. 1.3. Интерфейс программы TPF [Fonarev, Graphchikov, Konilov, 1991]. Для создания пользовательского интерфейса используется псевдографика. В программу загружены результаты микрозондового анализа состава минералов

принципов, лежащих в основе геобаротермометрии, существует большая опасность некорректного применения ее методов. Кроме рассмотренных выше требований к минеральной ассоциации, все калибровки сенсоров выполняются для конкретных горных пород и диапазонов содержания химических элементов. Некоторые геосенсоры требуют сложного много-ступенчатого расчета (см., например, в [Перчук, 2004]) или специфических методов подготовки исходных данных (например, метод предложенный в [Putirka, 2008]).

Следует также отметить, что ежегодно появляются новые публикации, в которых как публикуются новые методы или калибровки, так и подвергаются пересмотру уже опубликованные. Это создает ситуацию, при которой коллекция методов должна не только предоставлять возможность выполнять расчеты с помощью того или иного инструмента, но и постоянно обновляться с учетом новых опубликованных результатов.

1.6. Базы геохимических данных

Результаты геохимических анализов, как правило, публикуются и распространяются вместе с научными статьями, однако существуют крупные базы, предоставляющие доступ через Интернет к геохимическим данным, собранным из различных источников.

База *GEOROC* (Geochemistry of Rocks of the Oceans and Continents, <http://georoc.mpch-mainz.gwdg.de/georoc/>) является коллекцией опубликованных анализов вулканических горных пород и мантийных ксенолитов. Она содержит концентрации основных и рассеяных элементов, соотношения радиогенных и нерадиогенных изотопов, оценки возрастов для горных пород, вулканических стекол, минералов и включений. Образцы представляют 11 различных геологических обстановок. Метаданные включают, помимо прочего, географические координаты, характеристику горных пород, степень их измененности, метод проведения анализа, литературную ссылку и др. Данная база поддерживается Институтом химии им. Макса Планка города Майнц, Германия [*GEOROC database (Geochemistry of Rocks of the Oceans and Continents)*, 2004].

База содержит порядка миллиона результатов анализов для 443 тысяч образцов. Данные доступны для поиска по критериям или для загрузки в виде тематически организованных файлов. Файлы организованы по тематическому (архейские кратоны, внутриплитные вулканы и т.п.), и формальному (конкретные минералы и горные породы) принципам.

База *PetDB* (Petrological Database, <http://www.earthchem.org/petdbweb>) агрегирует данные о химическом и изотопном составе горных пород и минералов преимущественно по изверженным и метаморфическим породам дна океана, в особенности, базальтам срединно-океанических хребтов и абиссальным перидотитам, а также образцы ксенолитов мантии и нижней коры. *PetDB* поддерживается Геоабсерватория Ламонта-Доерти Колумбийского университета [*PetDB, the Petrological Database*, 2012].

Поиск по базе доступен по географическим координатам, типу горной породы, тектонической обстановке, типу образца, публикации, автору и типу данных.

База *NAVDAT* (The North American Volcanic and Intrusive Rock Database, <http://www.navdat.org/>) является коллекцией данных по возрастам,

химическим и изотопным составам Мезозойских и более молодых вулканических горных пород Северной Америки. Она создана коллаборацией ученых из четырех американских университетов и мексиканского Центра наук о Земле.

База содержит порядка 65 тысяч образцов. Поиск возможен по географическим координатам, региону, возрасту, составу, типу горных пород, публикации, ключевым словам, наличию определенных минералов. Также возможна загрузка базы целиком. База NAVDAT не обновляется с 2014 года [NAVDAT, *The North American Volcanic and Intrusive Rock Database*, 2014].

База LEPR (Library of Experimental Phase Relations, <http://lepr.ofm-research.org/>) агрегирует результаты опубликованных экспериментальных исследований, имеющих отношение к природным магматическим системам. База создана в рамках проекта OFM Research при поддержке Национального научного фонда США [*Library of Experimental Phase Relations*, 2008].

Также следует отметить существование крупного проекта *EarthChem* (<http://georoc.mpch-mainz.gwdg.de/>), индексирующего базы NAVDAT, SedDB [SedDB, 2013], PetDB, NavDat, GEOROC и Geochron [Geochron, 2011].

Представленный обзор баз геохимических данных не является исчерпывающим, однако позволяет сделать некоторые общие выводы относительно их организации. В частности, все подобные базы формируются по тематическому принципу, включая данные пород определенной группы или по определенной территории. При этом, данных по США существенно больше, чем по другим странам. Базы геохимических данных, как правило, имеют продвинутые инструменты поиска, но не имеют API для подключения сторонних приложений. Метаданные в таких базах включают, как минимум, географические координаты точки отбора образца, библиографическую ссылку на исходную публикацию и информацию о методе анализа. Данные выдаются пользователю для загрузки в формате простого текста с разделителями или электронных таблиц.

Пополнение большинства баз ведется на добровольных началах, например, на сайте *PetDB* находится образец таблицы Microsoft Excel, которую необходимо заполнить для добавления собственных данных в

базу. При этом требованием является предварительная публикация составов горных пород в реферируемом научном журнале.

1.7. Выводы

Программное обеспечение для работы с петролого-геохимическими данными представлено целым рядом инструментов от электронных таблиц до веб-приложений. Следует отметить, что программное обеспечение для обработки данных о составе минералов в целом представлено менее широко, чем инструменты для обработки валового состава горных пород.

Несмотря на разнообразие решений и широту их возможностей, следует отметить некоторые проблемы, общие для данного класса приложений. Первой из них является недостаточный объем справочной информации о методах обработки данных. В случае их небольшого числа, можно предположить, что пользователь, будучи специалистом, самостоятельно выберет нужный метод. Однако если число инструментов измеряется сотнями, то наличие справочной информации о методе становится важным условием *осознанного* выбора пользователем того или иного метода [Иванов, 2015b]. Кроме того, со временем в литературе накапливается информация, посвященная критической оценке различных методов. Некоторые из них признаются устаревшими, однако ни одна из рассмотренных программ не предоставляет информацию об этом. Хочется отметить, что подобная работа требует анализа огромного количества публикаций. Данная работа не может быть выполнена разработчиками ПО в одиночку, однако привлечение сообщества заинтересованных квалифицированных пользователей-геологов может позволить решить эту задачу.

Другой проблемой является закрытость исходных кодов и зависимость от платформы большинства существующих приложений. Использование в научных работах программного обеспечения с открытым исходным кодом (СПО) является залогом их воспроизводимости. Например, аргументы в пользу использования СПО в науке приводят представители редколлегии авторитетного научного журнала Nature [Ince, Hatton, Graham-Cumming, 2012]. Также, среди рассмотренных приложений практически нет кроссплатформенных, а некоторые из продуктов жестко привязаны к определенной коммерческой среде. Это не только затрудняет использование методов, заложенных в программе конечным пользователем, но, что

важнее, ограничивает возможность свободно делиться промежуточными результатами.

Третьей проблемой является невозможность расширения набора методов, имеющих в программах. Большинство рассмотренных приложений не предусматривают пополнения своих коллекций методов или такая возможность доступна только разработчикам. Следствием данной ситуации является то, что вместо совместного участия в расширении возможностей программного продукта потенциальные разработчики методов создают отдельные расчетные программы для своих нужд, ограниченные по функциональности и часто не распространяемые вовсе.

Четвертая проблема состоит в отсутствии возможности для пользователей делиться своими собственными наработками и рекомендациями по использованию методов. Ни одно из рассмотренных приложений не имеет социальных инструментов. Более того, у большинства продуктов отсутствуют официальные списки рассылки или форумы для общения пользователей.

Все это делает актуальным создание нового программного продукта лишенного рассмотренных выше недостатков. Для этого необходим качественно иной подход к архитектуре и дизайну приложения, который бы позволил не только разработать эффективный инструмент для обработки петролого-геохимических данных, но и предоставил бы пользователям возможность непосредственно участвовать в его развитии. Результатом поиска такого подхода стала концепция интерактивного реестра, рассмотренная в главе 2.

Глава 2. Интерактивный реестр методов обработки петролого-геохимических данных на основе технологии вики

В данной главе предложен отличный от существующих на сегодняшний день подход к построению приложения для обработки петролого-геохимических данных. В его основе лежит концепция информационной системы, объединяющей в себе документацию, исходный код методов, средства их использования. Кроме того система доступна для расширения самими пользователями. Она получила название *интерактивный реестр* [Иванов, 2016а]. Введение этого термина было обусловлено тем, что данная совокупность возможностей не покрывается в полной мере такими распространенными понятиями, как «база данных», «коллекция» или «библиотека», в то же время термин «информационная система» является слишком неопределенными. Также его введением хочется подчеркнуть наличие некоторых требований и правил, применяемых при описании методов в реестре.

2.1. Концепция интерактивного реестра

Основу концепции интерактивного реестра составляет ряд принципов, отличающих его от других приложений для обработки петролого-геохимических данных:

1. Информация о методе и его алгоритм неразрывны. В противовес системам, в которых дается текстовое или формальное описание метода (например, в тексте статьи), либо содержится реализация алгоритма метода в виде, недоступном для восприятия пользователем (например, программы расчета в электронных таблицах), в реестре информация о методе и его алгоритм должны размещаться совместно и быть в равной степени доступны.

2. Описание алгоритма на языке программирования непосредственно используется для вычислений. Использование алгоритма возможно вне реестра. Система должна быть разработана так, чтобы избежать расхождения между алгоритмом, приведенным в описании метода, и

алгоритмом, использованным при расчете. Естественным решением данной проблемы является использование одного и того же исходного кода для описания методов и выполнения расчетов. Этот подход предъявляет определенные требования к языку программирования как способу описания алгоритма метода. В частности, язык должен содержать минимальное количество несущественных для алгоритма синтаксических конструкций и быть легким для восприятия людей, мало знакомых с программированием. Кроме того, использование метода должно быть доступно без слишком больших усилий вне реестра в виде библиотеки или подключаемого модуля. Это требование также налагает ограничения на выбор языка.

3. Информация о методах организована. Реестр должен являться не только вычислительным инструментом, но и быть организованным по принципу справочника. Принципами такой организации являются: группировка по категориям, добавление ключевых слов, использование перекрестных и библиографических ссылок, возможности поиска. Реестр должен обеспечивать возможность работы с сотнями методов, поэтому вопрос организации информации является существенным для его построения.

4. Информация о методах, в том числе их алгоритмы, могут развиваться. Со временем не только количество методов, но и объем информации о них будет расти, соответственно реестр также должен развиваться. Основными направлениями такого развития являются: пополнение реестра новыми методами, совершенствование документации к существующим (в том числе добавление примеров использования), обсуждение и историческое ранжирование, а также исправление обнаруженных ошибок. Для успешного решения этих задач реестр должен содержать инструменты совместной работы, управления версиями, рецензирования и т.д.

5. Реестр доступен для внесения изменений пользователями. Концепция интерактивного реестра подразумевает не только его использование, но и модификацию пользователями. Это может быть пополнение реестра новыми методами, дополнение информацией о них, их обсуждение. Привлечение пользователей, в том числе не программистов к работе по развитию реестра является новой идеей для данного класса приложений. Как правило, в разработке участвует только небольшая группа авторов — геологов и программистов, однако при таком большом числе методов обработки, которое имеется, например, в области геобаротермометрии,

даже небольшой вклад от большого числа пользователей может принести существенную пользу проекту.

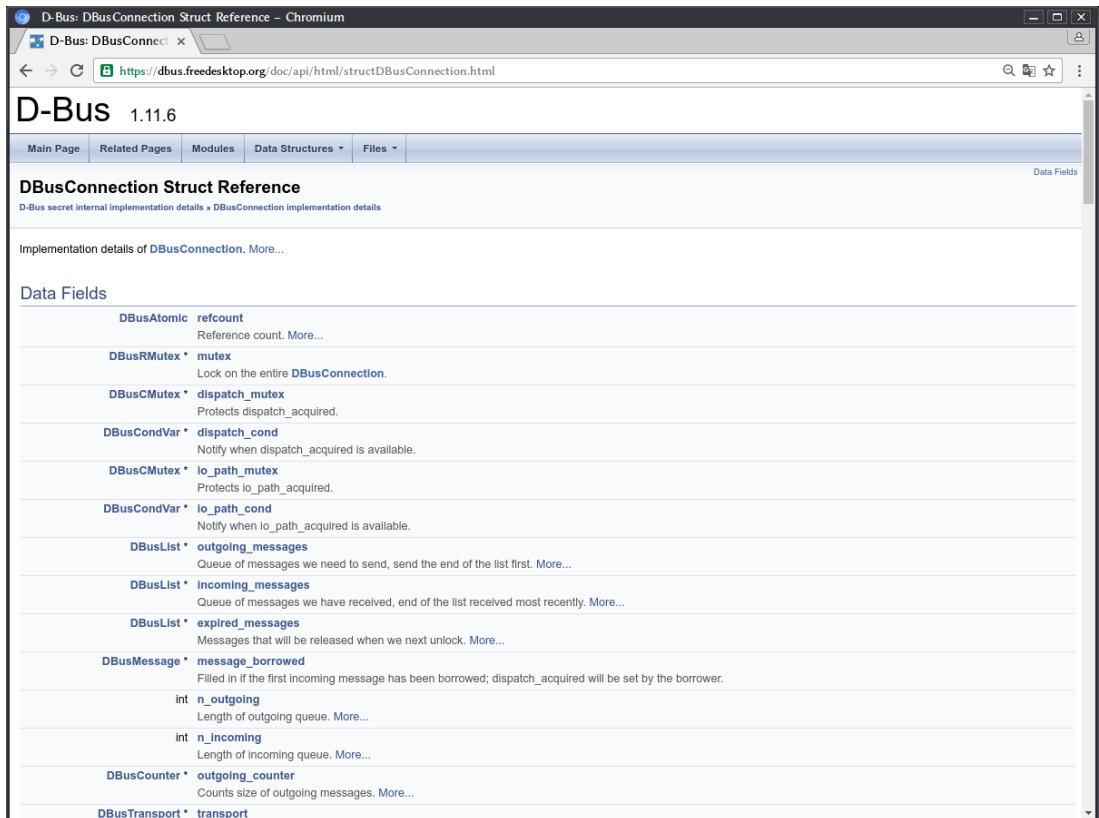
Для создания приложения в соответствии с этими принципами необходимо решить задачи организации и управления формализуемой и неформализуемой информацией о методах, а также организации вычислений.

2.2. Интерактивный реестр на основе веб-приложения

Для реализации интерактивно реестра необходимо выбрать архитектуру, которая позволила бы построить приложение в соответствии с принципами интерактивного реестра. Ключевым аспектом при выборе архитектуры являются отношения между исходным кодом метода и его описанием. Традиционный подход к построению библиотек методов, сохранившийся, например, в библиотеке алгоритмов LAPACK [Anderson et al., 1999], состоит в ведении отдельной от кода документации. Данный подход сложился в те времена, когда документация распространялась в виде печатной продукции, как, например, руководство программиста операционной системы Unix [Laboratories, Kernighan, McIlroy, 1979]. Существенной проблемой такого подхода является необходимость дублирования информации в виде кода и в виде документации. Кроме того, трудно контролировать отсутствие разночтений между кодом библиотеки и ее документацией. Для решения этой проблемы был предложен подход, при котором документация включается непосредственно в исходный код в виде комментариев. Основное преимущество этого подхода состоит в отсутствии дублирования информации, так как генератор документации анализирует не только комментарии, но и непосредственно исходный код. Примером такого подхода являются системы документирования исходного кода Javadoc [*javadoc - The Java API Documentation Generator*, 2004] и Doxygen [Heesch, 1997] (см. рис. 2.1). Аналогичные технологии используются в некоторых пакетных системах, например в библиотеке модулей для платформы R — CRAN и CTAN для издательской системы Tex.

Рассмотренные выше подходы базируются на нескольких допущениях, вполне справедливых для большинства случаев разработки программного обеспечения. Первое допущение состоит в том, что авторы самого метода,

a.



б.

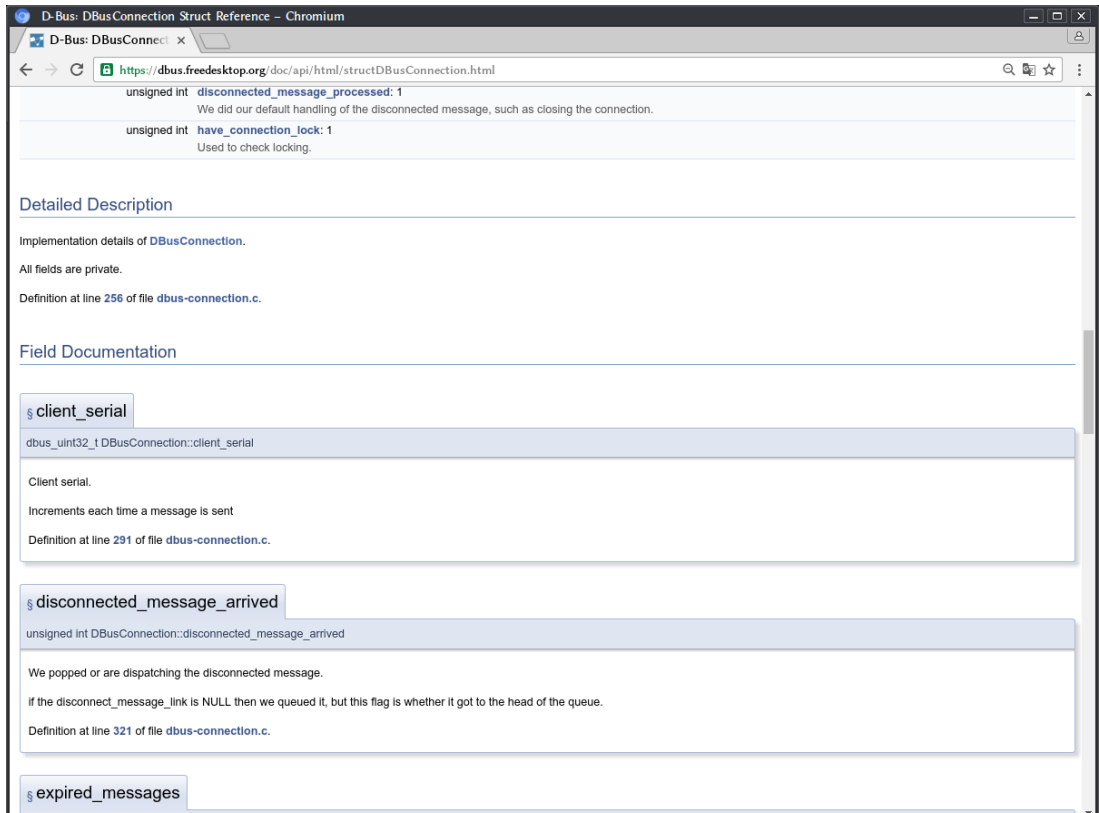


Рис. 2.1. Пример документация, сгенерированной системой Doxygen для библиотеки D-BUS (<https://dbus.freedesktop.org/>). Страница с описанием структуры данных (а) и подробное описание полей (б)

его исходного кода и документации если и не являются одним и тем же человеком, то, по крайней, мере представляют одну организацию, в рамках которой можно ввести определенные правила написания кода и оформления документации. Второе допущение предполагает, что вся необходимая для использования метода информация доступна к моменту его реализации.

В рамках концепции интерактивного реестра данные допущения не являются верными. Автор реализации метода на языке программирования, автор самого метода и автор описания метода скорее всего являются разными людьми, возможно никак не взаимодействующими между собой. По мере публикации новых работ, посвященных тем или иным методам, документация к ним может развиваться. Кроме того, развитие документации может быть существенно более глубоким процессом, чем развитие реализации метода, которая может быть достаточно проста. Общение между разработчиками и пользователями происходит не только по специальным каналам, таким, как списки рассылки и форумы. Пользователи непосредственно привлекаются к участию в проекте. Соответственно, допущение, что авторы документации будут в точности следовать разработанным правилам, неверно.

В этих условиях логично было бы обратиться к системам коллективной разработки. Такие системы представлены на сегодняшний день решениями двух типов. Первым типом являются системы коллективной разработки программного обеспечения. В простейшем случае это системы управления версиями (VCS). Они позволяют создавать централизованное или децентрализованное хранилище файлов с поддержкой контроля версий и механизмом разрешения конфликтов, предназначенное для совместной работы над исходным кодом. Примерами таких систем являются SVN [Collins-Sussman, Fitzpatrick, Pilato, 2004], Git [Loeliger, McCullough, 2012], Bazar [Gyerik, 2013] и другие. Также на их базе могут быть созданы интегрированные решения для разработки ПО, такие как Stash <https://www.atlassian.com/software/bitbucket/server> или GitHub <https://github.com>. Следует отметить, что эти системы в первую очередь, предназначены для работы с собственно исходным кодом, а не с документацией. Вторым типом являются системы коллективной работы над документами, в частности, вики-системы [Leuf, Cunningham, 2001], подробнее рассмотренные в разделе 2.2.1.

2.2.1. Технология вики

В середине 90-х годов двадцатого века в классе систем управления содержанием веб-сайтов (CMS) выделился новый тип. В отличие от традиционных CMS, имеющих отдельные интерфейсы для пользователей и редакторов, они имели единый интерфейс редактирования, доступный всем пользователям, возможно, с определенными ограничениями. Они были ориентированы на привлечение пользователей к созданию материалов сайта. Эти системы получили название вики [Leuf, Cunningham, 2001]. Сайты на основе вики являются представителями технологии Web 2.0 [O'reilly, 2007], в рамках которой граница между создателями и потребителями информации постепенно стирается (см. рис. 2.2). Вики отличаются от коллективных блогов возможностью нескольким авторам совместно работать над одними и теми же материалами. Наиболее широко известным проектом, использующим эту технологию, является Википедия (<http://wikipedia.org>). В сфере наук о Земле в России известен вики-проект «Все о геологии» (<http://geo.web.ru/>).

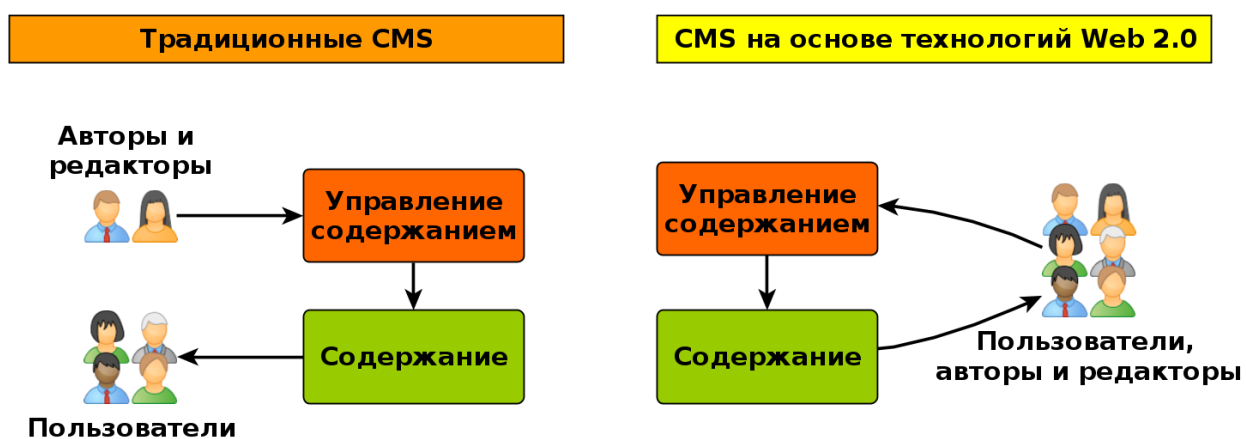


Рис. 2.2. Два различных подхода к организации систем управления содержанием (CMS). В рамках традиционных CMS (слева) существует жесткое разделение на производителей и потребителей информации. В CMS на основе технологий Web 2.0 (справа) производители и потребители информации рассматриваются как единое сообщество пользователей системы

Чертами вики-систем, интересными в рамках данной работы, являются:

- Возможность многократно редактировать материалы сайта с помощью встроенных в него инструментов без применения стороннего программного обеспечения;
- Контроль версий материалов: возможность сравнения редакций, просмотра и восстановления ранних версий;
- Специальный язык разметки — так называемая вики-разметка, позволяющая сравнительно легко форматировать и оформлять отдельные структурные элементы материала;
- Пространства имен — механизм иерархической организации содержимого вики.

Эти черты позволяют использовать систему на основе вики для хранения и организации всей информации о методах в рамках реестра.

В качестве основных недостатков систем на основе этой технологии можно назвать необходимость освоения специального языка разметки. Общей проблемой всех систем на базе Web 2.0 считаются те или иные формы вандализма. Однако современные вики-системы могут быть как публичными, так и иметь различные ограничения, в частности, ограничение на внесение изменений, что требует регистрации авторов и редакторов, а также, возможно, модерирования материалов.

2.2.2. Выбор вики-платформы

Поскольку создание новой вики-системы выходит за рамки цели и задач данной работы, предлагается использовать одну из существующих систем в качестве платформы для построения реестра. Платформе предлагается передать следующие общие функции:

- хранение и представление материалов — описаний и алгоритмов методов;
- редактирование материалов и управление их версиями;
- навигация и поиск по материалам;
- авторизация и управление пользователями.

Также выбранная вики-система будет использоваться в качестве основы для построения графического пользовательского интерфейса.

Существующие на сегодняшний день вики представлены как самостоятельными CMS (например, MediaWiki <https://www.mediawiki.org/>,

DokuWiki <https://www.dokuwiki.org/>, MoinMoin <https://moinmo.in/> и др.), так и компонентами в составе крупных систем управления корпоративными ресурсами и документооборота (например, Alfresco <https://www.alfresco.com/>) и служб учета сообщений об ошибках (например, Trac <https://trac.edgewall.org/>). Кроме того, существуют интегрированные пакеты приложений на основе технологии вики (например, FosWiki <https://foswiki.org/>).

Для выбора конкретной системы был выработан ряд критериев, в частности, из рассмотрения были исключены несвободные программные продукты, так как помимо лицензионных трудностей для успешной реализации реестра на определенном этапе может потребоваться модификация самой вики-системы, что в случае несвободных программных продуктов не всегда возможно. Также требуется, чтобы вики-система не была частью более крупного продукта, возможности которого не будут использованы в реестре. С другой стороны, она должна иметь интерфейс для подключения модулей расширения (плагинов), для обеспечения вычислений. Кроме того, дополнительными критериями стали активное развитие проекта, наличие крупного сообщества разработчиков. Эти факторы обеспечивают поддержку и обновления для вики-системы в будущем.

Среди удовлетворяющих этим требованиям CMS на основе технологии вики можно выделить системы MediaWiki [Barrett, 2009] и DokuWiki [Victorino, 2012]. Обе являются самостоятельными CMS, свободным программным обеспечением (распространяются под лицензией GNU GPL), имеют интерфейс для подключения модулей, обладают богатым набором возможностей по организации информации и имеют активное сообщество разработчиков. С технической точки зрения обе системы написаны на языке программирования PHP и используют достаточно схожие языки вики-разметки, поддерживают управление версиями, позволяют загружать изображения и другие файлы. Основные различия между этими системами отражены в таблице 2.1.

MediaWiki разработана для крупных проектов, рассчитанных на очень большое количество пользователей (во многом его внутренняя архитектура обусловлена нуждами Википедии — одного из самых посещаемых сайтов в сети Интернет, который использует эту CMS), в то время, как DokuWiki

Таблица 2.1. Сравнение CMS MediaWiki и DokuWiki

Критерий	MediaWiki	DokuWiki
Хранилище	База данных	Файловая система
Защита от конфликтов правок	Блокировка страницы	Уведомление
Контроль доступа	Режим «Только чтение»	ACL
Организация страниц	«Плоская»	Иерархическая

предназначена для небольших проектов. Эти различия определяют особенности их реализации. Для хранения материалов MediaWiki использует базу данных, в то время как DokuWiki хранит их в текстовых файлах. Механизм блокировок при совместном редактировании страниц в DokuWiki очень примитивен, т.к. не рассчитан на большой поток правок. MediaWiki ориентирована на создание публичных вики-систем, поэтому возможности по разграничению доступа к ней невелики, в то время как DokuWiki может использовать мощные инструменты управления пользователями, такие как списки контроля доступа — ACL. Несмотря на некоторые недостатки, в виду своей простоты и расширяемости, в качестве платформы для построения интерактивного реестра была выбрана система DokuWiki.

2.2.3. Организация информации о методах в рамках вики

В ходе анализа предметной области — методов обработки петролого-геохимических данных — были выделены некоторые категории методов, которые характеризуются однородными входными и выходными данными [Иванов, 2016а]. Механизм пространств имен в DokuWiki аналогичен по своей работе (и внутреннему представлению) механизму папок в файловой системе. В качестве символа разделителя в пространстве имен используется двоеточие. Кроме того, для каждого пространства имен существует специальная индексная страница. Для каждой категории методов было выделено отдельное пространство имен, а общая информация была размещена в корневом пространстве. При этом структура реестра приобрела вид показанный на рисунке 2.3.

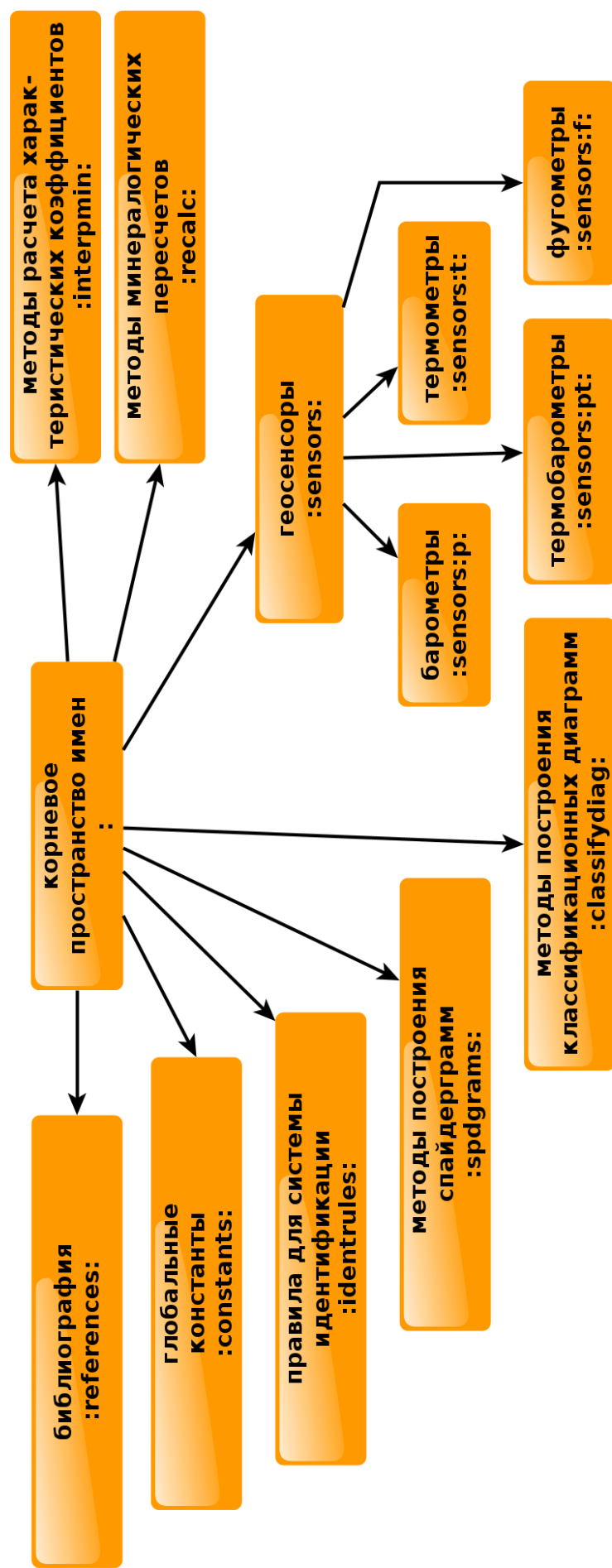


Рис. 2.3. Структура интерактивного реестра. Показаны соотношения между основными пространствами имен

Были добавлены специальные пространства имен для организации библиографической информации (:reference, см. раздел 2.2.4) и модулей пользовательского интерфейса для выполнения расчетов (:ppu, см. раздел 2.3.7).

Единицей хранения информации в CMS DokuWiki является страница, имеющая уникальное имя. Для хранения каждого метода в рамках вики-системы использоваться отдельная страница. Имя страницы используется в ее URL адресе. Таким образом, адрес страницы метода в реестре будет состоять из названия пространства имен и названия страницы. В каждом пространстве имен на индексной странице был размещен список всех методов (то есть страниц) в нем. Для этого был использован плагин к системе DokuWiki под названием nspages. Страницы вики автоматически индексируются системой, что обеспечивает возможность поиска по ним с помощью встроенного инструмента. Кроме того, за каждой страницей с помощью плагина tag может быть закреплён один или несколько тегов для дальнейшей организации.

Для размещения разнородной информации о методе на одной странице используются элементы вики-разметки. Для всех страниц методов используется общий шаблон, представленный на листинге 2.1.

Для формирования заголовка страницы могут использоваться различные правила для разных категорий методов. Имя страницы должно соответствовать заголовку с учетом технических ограничений, налагаемых на URL адреса (только латиница, замена пробелов и знаков пунктуации на символ нижнего подчеркивания). Например, страница, посвященная амфибол-плагиоклазовому сенсору Холонда и Бланди [Holland, Blundy, 1994], имеет заголовок «Amp-Pl: Holland, Blundy, 1994» и полное имя страницы `sensors:p:amp_pl_holland_blundy`.

```

1 ===== [Заголовок страницы] =====
2
3 [Краткое описание метода]
4
5 === [Подробнее] ===
6
7 ^ [Ссылка] | [Библиографическая ссылка] |
8 ^ [Поле] | [Значение] |
9 ^ [Поле] | [Значение] |
10
11 [Ссылки для использования данного метода]
12
13 [Подробное описание метода]
14
15 [Информация о состоянии страницы]
16
17 [Дополнительная информация о методе]
18
19 === [Исходный код] ===
20
21 <Code>
22 [Исходный код метода]
23 </Code>

```

Лист. 2.1. Шаблон страницы метода в интерактивном реестре. Текст в квадратных скобках означает подстановку соответствующего значения в данное поле

В таблице [Поле] | [Значение] листинга задаются формальные характеристики для данного типа методов, такие как границы входных и выходных параметров. Эти поля могут быть использованы при автоматической обработке страницы. Текст из поля [Краткое описание метода] будет отображаться на странице выбора метода в качестве краткого описания. Шаблон не ограничивает возможность добавления других разделов, например, с замечаниями по использованию данного метода. В поле [Библиографическая ссылка] вставляется специальная ссылка на страницу из раздела библиографической информации (см. раздел 2.2.4).

2.2.4. Организация работы с библиографией в рамках вики

При создании реестра невозможно обойтись без работы с библиографией в том или ином виде. Библиографическая ссылка является первой и необходимой информационной записью о любом методе. Для эффективной

работы с библиографической информацией было принято решение организовать ссылки на статьи в виде страниц в отдельном пространстве имен `:references:`. А для хранения собственно библиографических записей использовать распространенный формат BibTeX.

Для добавления библиографической записи необходимо добавить новую страницу в упомянутом пространстве имен и оформить ее, используя специальный шаблон (см. листинг 2.2). При этом имя страницы должно, в целях унификации, содержать фамилию первого автора и год публикации. В блоке `code` шаблона необходимо разместить библиографическую запись в формате BibTeX.

```
1 ===== [Название статьи, автор, год] =====
2
3 <fullref:[Имя статьи]>
4
5 [Внешняя ссылка на статью]
6
7 <code>
8 [Библиографическая запись в формате BibTeX]
9 </code>
```

Лист. 2.2. Шаблон страницы библиографической информации в интерактивном реестре

После сохранения страницы на месте конструкций `<fullref:[название_страницы]>` и `<ref:[название_страницы]>` появятся подробная и краткая библиографические ссылки соответственно. Данный механизм реализован с помощью разработанного автором плагина к системе Dokuwiki, который выполняет разбор BibTeX записи и формирует эти элементы страниц. Данные элементы расширяют стандартную вики-разметку. Их можно использовать на любых страницах реестра, в частности, на страницах соответствующих методов как в качестве основной ссылки на статью, так и в качестве дополнительных ссылок.

Кроме этого, на страницу в пространстве `:references:` можно добавить внешнюю ссылку на полный текст статьи (например, ссылку через систему DOI или непосредственно URL на сайте журнала), с помощью следующей команды: `<filelink:[URL]|[Имя файла]>`. Следует отметить,

что для большинства статей доступ к полному тексту через Интернет возможен только по подписке или за отдельную плату.

2.3. Методы обработки геохимических данных в интерактивном реестре

Вики-система предоставляет многочисленные инструменты для организации информации, однако не содержит встроенных компонентов для выполнения каких либо расчетов, поэтому для реализации реестра необходимо в дополнение к ней разработать и реализовать инструменты выполнения методов, которые, в соответствии с концепцией реестра, должны размещаться в виде исходного кода непосредственно на вики-странице.

Кроме рассмотренной в разделе 2.2.4 автоматической обработки библиографических записей, следующие функции должны быть реализованы в дополнение к возможностям вики:

- Загрузка и управление собственными данными пользователей;
- Использование имеющиеся в реестре методы для выполнения расчетов.

Для выполнения этих функций автором был разработан на языке Python3 специальный промежуточный модульный сервер. При его разработке использовался набор шаблонов проектирования модель-представление-контроллер (MVC) [Reenskaug, 2003]. В соответствии со сложившимися практиками разработки веб-приложений, представление в браузере элементов пользовательского интерфейса выполняется средствами языков HTML и CSS. В качестве контроллера выступает набор скриптов, написанный на JavaScript, также выполняемый в браузере. В качестве модели выступает интерфейс промежуточного сервера, разработанный в соответствии с принципами REST [Fielding, 2000]. Взаимодействие контроллера с моделью происходит через выполнение HTTP запросов в асинхронном режиме с помощью AJAX. Также был разработан модуль к системе DokuWiki, который обеспечивает бесшовную интеграцию интерфейса выполнения расчетов с реестром. В качестве файлового менеджера был использован свободный файловый менеджер elFinder. Общая архитектура реестра представлена на рисунке 2.4.

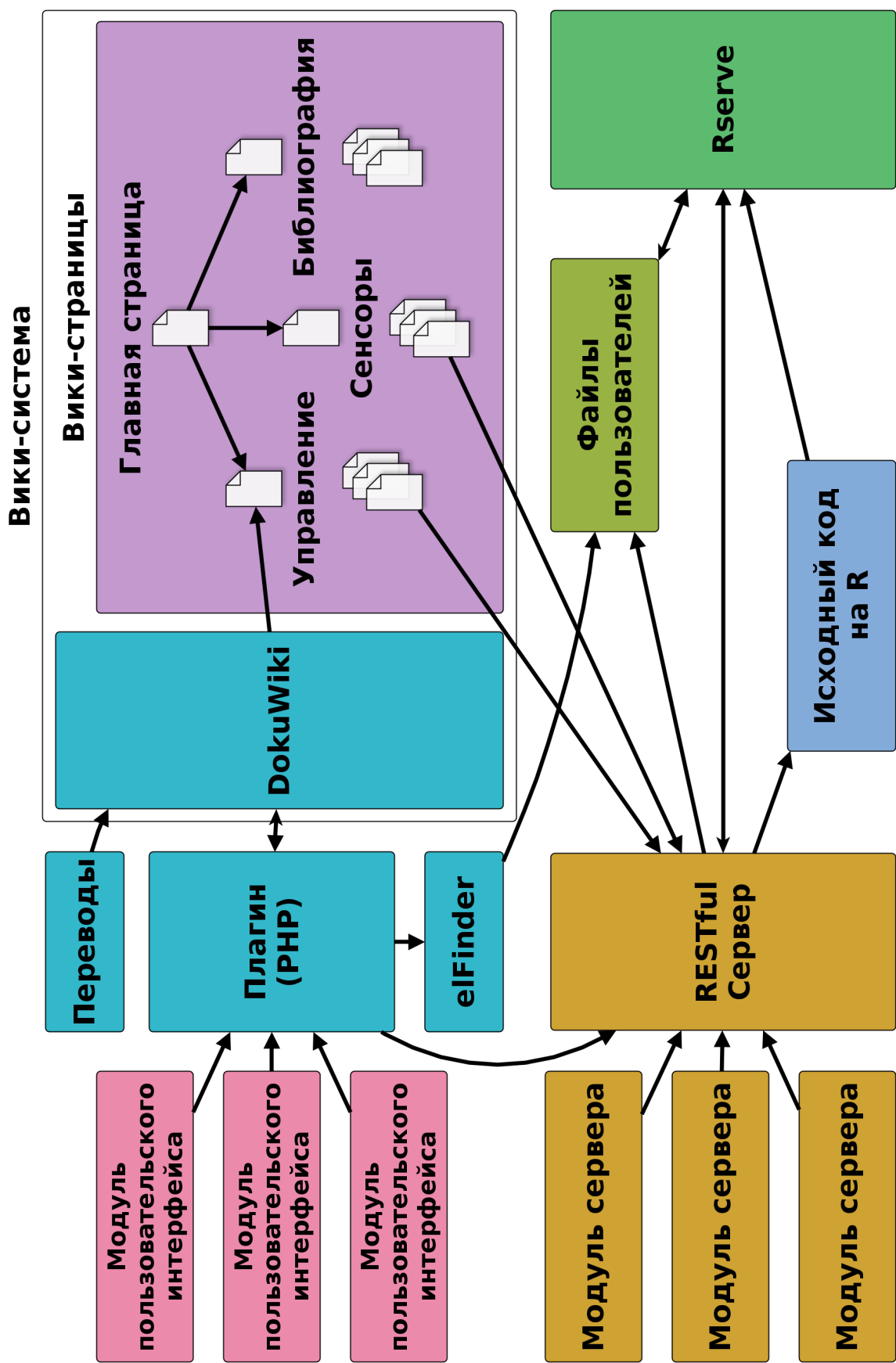


Рис. 2.4. Программная архитектура интерактивного реестра. На диаграмме показаны основные компоненты: вики-система, плагин к ней, промежуточный RESTful сервер и сервер вычислений Rserve

HTTP API, сервера имеет префикс /ppu и содержит несколько веток:
/ppu/data/* — управление, просмотр и преобразование данных пользователя, в том числе импорт и экспорт;
/ppu/viz/* — вспомогательные методы визуализации;
/ppu/rpin.zip — выгрузка библиотеки методов. (см. раздел 2.5);

Сервер использует модифицированную версию библиотеки web.py [Swartz, 2016] для обслуживания HTTP запросов. Эта библиотека реализует необходимые примитивы для разбора HTTP заголовков и обеспечивает многопоточность при обработке запросов. Данные сервером и браузером передаются в формате JSON.

Учитывая существование различных категорий методов обработки геохимических данных, сервер был сделан модульным. Каждый модуль промежуточного сервера отвечает за работу с конкретной категорией методов и имеет отдельную ветку в API, например, /ppu/recalc/* для минералогических пересчетов. Также сервер предоставляет каждому модулю набор примитивов для выполнения вычислений и обработки запросов.

Промежуточный сервер выполняет проверку наличия у пользователя открытой сессии DokuWiki. Таким образом, только пользователь, вошедший в вики-систему, получает доступ к выполнению запросов к промежуточному серверу, в том числе, к выполнению расчетов. Кроме того, сервер с помощью механизма inotify [Kroah-Hartman, 2006] получает информацию об обновлении страниц вики. Промежуточный сервер работает в режиме службы и получает ключевые настройки, такие как путь к DokuWiki и порт из конфигурационного файла.

Для запуска разработанной системы, состоящей из DokuWiki и промежуточного сервера, требуется выделенный сервер (возможно виртуальный) под управлением ОС GNU/Linux. На нем должны быть установлены актуальные версии Apache, PHP и Python3. Для совместной работы HTTP сервера Apache и промежуточного сервера, первый должен быть настроен в режиме обратного проксирования запросов по адресам, начинающимся с /ppu на порт промежуточного сервера. Также возможна работа по протоколу HTTPS, для чего необходимо получить соответствующий сертификат и выполнить настройку сервера Apache.

2.3.1. Управление пользовательскими данными

Наиболее распространенным типом геохимических данных является таблица, содержащая результаты химических измерений для различных образцов. Результатами работы методов также являются таблицы или различные изображения (см. раздел 2.4). Таким образом, реестр должен предоставлять пользователю интерфейс для загрузки, выгрузки и просмотра таблиц и изображений. Разнотипность этих данных наводит на мысль об использовании файлового менеджера для работы с ними. В рамках этого подхода каждому пользователю выделяется собственная папка для хранения файлов, а все операции сводятся к управлению файлами в этой папке. Соответственно, необходим обозреватель, предоставляющий пользователю возможности по управлению файлами. Система DokuWiki не содержит подобных инструментов. Поэтому для управления файлами пользователя в вики с помощью специально разработанного автором плагина был интегрирован свободный файловый менеджер eFinder (рис. 2.5).

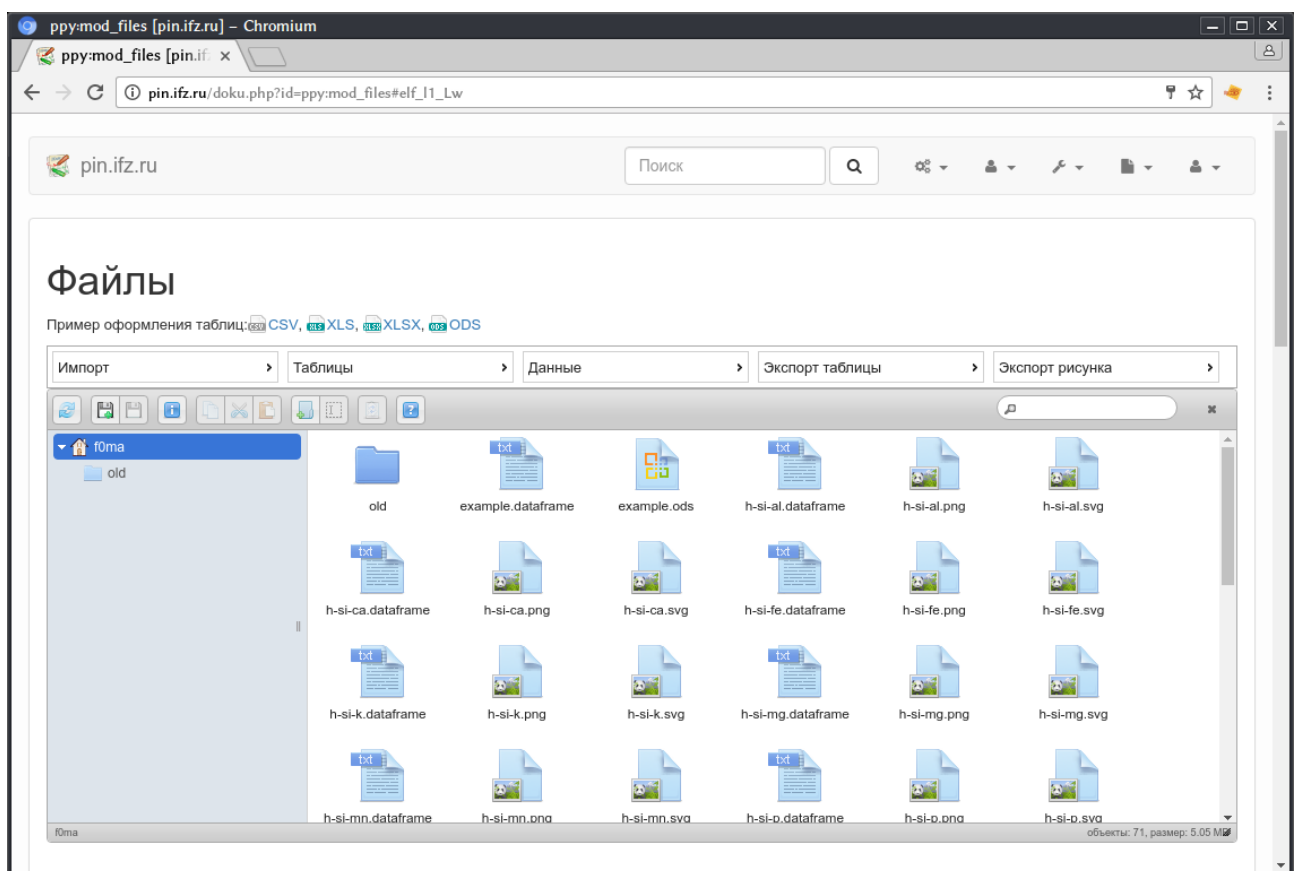


Рис. 2.5. Файловый менеджер eFinder (<https://studio-42.github.io/eFinder>), интегрированный с DokuWiki. Выше окна файлового менеджера расположено меню основных операций над файлами (см. прил. 1)

На файловый менеджер и разработанный плагин были возложены следующие функции:

- Загрузка и выгрузка пользовательских файлов;
- Базовые операции (создание папок, переименование, удаление, сортировка файлов);
- Просмотр таблиц и изображений;
- Импорт и экспорт таблиц;
- Преобразование типов таблиц.

Первые две из этих функций уже реализованы в elFinder, однако импорт, экспорт таблиц, их просмотр и преобразование были добавлены через интерфейс расширений, который предоставляет данный файловый менеджер. Импорт необходим, так как типичные форматы файлов электронных таблиц (XLS, XLSX, ODS) не предусматривают достаточную формализацию данных для использования в методах, поэтому необходимо их преобразование в простой и достаточно формальный вид. В качестве такого формата был выбран формат CSV в соответствии с RFC4180 [Shafranovich, 2005]. Он обеспечивает простоту хранения табличных данных и удобство их дальнейшей обработки штатными инструментами языка R. В файловом менеджере импортированные файлы получают расширение `.dataframe` и становятся доступны в модулях выполнения расчетов. При импорте проверяется соответствие названия столбцов принятым для методов в реестре (см. раздел 2.3.3), при необходимости столбцы автоматически переименовываются.

Непосредственно преобразование форматов файлов и подготовка просмотра таблиц возложена на специальный модуль сервера, рассмотренного в разделе 2.3. Функции этого модуля доступны по адресам `/ppu/data/*` и включают, помимо преобразования файлов, функции для сохранения результатов вычислений, загрузки библиотеки методов в виде архива, разделения и объединения таблиц, а также преобразования форматов изображений из используемого в реестре SVG [Eisenberg, Bellamy-Royds, 2014] в другие векторные и растровые форматы (с использованием консольной версии векторного графического редактора Inkscape <https://inkscape.org/>).

Каждому пользователю в системе выделяется отдельный каталог для рабочих файлов, который рассматривается файловым менеджером как

корневой. В этот каталог пользователь может загружать исходные файлы и в него сохраняются, полученные им результаты. Пользователь может создавать подкаталоги для организации своих данных. Возможно создание администратором реестра средствами ОС ссылок на каталоги, общие для нескольких пользователей, что позволяет рабочим группам пользоваться общими исходными данными и делиться результатами.

2.3.2. Выбор языка программирования для описания методов

Для реализации в промежуточном сервере механизма выполнения методов необходимо выбрать язык, на котором будут записываться их алгоритмы в интерактивном реестре. Среди критериев выбора языка для описания методов можно выделить две группы. К первой относятся общие критерии, обусловленные принципами построения реестра. К ней можно отнести переносимость, распространенность, хорошую документацию и уверенность в том, что поддержка этого языка будет продолжаться как для существующих, так и для будущих программно-аппаратных платформ. Язык должен обладать простым синтаксисом и возможностью встраивания в сторонние приложения для эффективной интеграции с реестром. Более предпочтительными являются интерпретируемые языки, так как они, как правило, позволяют динамически загружать и перезагружать части исходного кода, которыми станут алгоритмы методов содержащиеся в реестре. Кроме того, язык должен быть свободным программным обеспечением, чтобы защитить реестр от изменения лицензионной политики разработчика.

Ко второй категории относятся критерии, обусловленные стоящей задачей — обработкой петролого-геохимических данных. К ним относятся, прежде всего, удобство работы с табличными данными, так как именно в этом виде представлены геохимические данные и простота организации вычислений. Методы обработки петролого-геохимических данных редко представляют собой ресурсоемкие задачи, поэтому производительность не является ключевым критерием при выборе языка.

В качестве возможных языков программирования из списка популярных языков [*The TIOBE Programming Community index*, 2016] можно выбрать ориентированные на вычисления MATLAB и R, а также язык общего назначения Python. Во всех этих языках работа с табличными данными

Таблица 2.2. Сравнение некоторых синтаксических элементов языков C, R и Python

C	R	Python
Объявление переменной и присвоение значения		
<code>float a = 2.5;</code>	<code>a <- 2.5;</code>	<code>a = 2.5</code>
Условный оператор		
<code>if(a > 0){ a = 42; }</code>	<code>if(a > 0){ a <- 42; }</code>	<code>if a > 0: a = 42</code>
Объявление функции		
<code>float f (float a){ return a*2+2; }</code>	<code>f <- function (a){ return(a*2+2); }</code>	<code>def f (a): return a*2+2</code>

достаточно удобна, однако MATLAB является коммерческим продуктом, что накладывает ограничения на его использование и встраивание (тем не менее, по крайней мере одна коллекция методов обработки геохимических данных для MATLAB существует [Dachs, 1998]). Синтаксис Python существенно отличается от привычных многим потенциальным пользователям C-подобных языков, в частности, использованием отступов для организации блоков кода (см. табл. 2.2), поэтому выбор был сделан в пользу R. Следует отметить, что для него также существует как минимум одна коллекция методов обработки геохимических данных GCDKit [Janousek, 2006].

Встраивание языка R может выполняться с помощью сервера вычислений RServe [Urbanek, 2003]. Данный сервер взаимодействует с клиентами через собственный протокол (по умолчанию используется 6311 TCP порт). Этот сервер позволяет выполнять вычисления и передавать данные аналогично командному интерпретатору R CMD. Существует реализация клиента протокола для языка Python под названием pyRserve, обеспечивающая прозрачный для программиста интерфейс вызова функций и передачи параметров между интерпретатором Python и сервером pyRserve (как локальным, так и удаленным). Таким образом, для выполнения расчетов в интерактивном реестре промежуточный сервер получает исходный код методов со страниц вики-системы и выполняет его с использованием Rserve.

2.3.3. Программный интерфейс методов

Язык R обеспечивает поддержку структурного программирования [Kabacoff, 2015], следовательно, самым рациональным способом взаимодействия с ним является вызов функций. Таким образом, программный интерфейс каждого из методов реестра должен состоять из одной или нескольких функций, вызов которых с передачей данных в качестве аргументов обеспечит его использование. Для этого необходимо создать конвенцию по именованию функций, которые бы обеспечили отсутствие неоднозначности. В интерактивном реестре имя функции соответствует имени страницы и пространству имен с условием, что символ двоеточия заменяются на символ точки допустимый в именах функций в R. Например, имя функции, реализующей метод на странице `sensors:p:amp_pl_holland_blundy`, записывается как `sensors.p.amp_pl_holland_blundy`. Также необходимо определить способ передачи аргументов, которые могут быть различны для разных методов. Для этого задействуется возможность передачи неопределенного списка аргументов, имеющаяся в R.

Рассмотрим шаблон исходного кода геотермометра (листинг 2.3). Помимо основной функции, в нем объявлено еще две вспомогательных. Функция `args` возвращает информацию об аргументах функции (в данном случае — список минералов и дополнительных параметров), а `test` — выполняет проверку правильности работы функции на тестовых данных. Тип, количество аргументов функций и их возвращаемое значение может быть различно в зависимости от того, к какой категории методов она относится. Например, геотермометр принимает составы минералов как таблицы встроенного типа `dataframe`, и значения свободных параметров как числа с плавающей запятой. Он возвращает таблицу результатов в формате `dataframe`. Формальная проверка работы метода на тестовых данных не гарантирует корректности его реализации, однако обеспечивает относительно простой и мощный механизм борьбы с возможными ошибками при реализации метода. Получение тестовых данных (т.е. составов минералов и значений свободных параметров в случае геосенсора) не всегда является тривиальной задачей. Часто данные, представленные в оригинальной статье, не полны (например, представлены только

пересчитанные данные). В этом случае требуется искать дополнительные источники контрольных примеров для формальной проверки корректности реализации метода.

```
1 sensors.t.[Имя сенсора] <- function([Список минералов],
2                                     [Список параметров], ...){
3   [Исходный код метода]
4   return([Возвращаемое значение]);
5 }
6
7 sensors.t.[Имя сенсора].args <- function(){
8   ars <- list();
9   ars$minerals <- [Список минералов];
10  ars$ranges <- [Список параметров];
11  return(ars);
12 }
13
14 sensors.t.[Имя сенсора].test <- function(){
15   result <- sensors.t.[Имя сенсора]([Тестовые данные]);
16   expect_equal(result, [Эталонный результат]);
17 }
```

Лист. 2.3. Шаблон исходного кода геотермометра

2.3.4. Правила представления данных для их однозначной интерпретации

Для правильной работы метода автором была разработана конвенция об именовании полей в таблицах. В соответствии с ней в каждой таблице (файле формата dataframe) первым столбцом должен идти столбец Name, содержащий имена образцов. Порядок остальных столбцов является произвольным.

Столбцы, содержащие геохимические данные, обозначаются именем соединения или элемента, а затем через нижнее подчеркивание указывается формат представления. Например, для весовых процентов оксидов — SiO2_wtp, TiO2_wtp и т.д., для атомных процентов Si_wtp, Ti_wtp и т.д., а для коэффициентов формул Si_pfu, Ti_pfu и т.д. В случае различных координаций или степеней окисления используется дополнительный инфикс, например Si_c4_pfu/Si_c6_pfu или Fe_p2_pfu/Fe_p3_pfu. Примеры записи названий столбцов в такой нотации представлены в таблице 2.3.

Таблица 2.3. Примеры нотации названий столбцов

Элемент	Si	Al	Fe
Весовые проценты содержания оксида	SiO2_wtp	Al2O3_wtp	FeO_wtp, Fe2O3_wtp
Весовые проценты содержания элементов	Si_wtp	Al_wtp	Fe_wtp
Коэффициент формулы для элемента	Si_pfu	Al_pfu	Fe_pfu
Коэффициент формулы для элемента в различных координациях	Si_c4_pfu, Si_c6_pfu	Al_c4_pfu, Al_c6_pfu	-
Коэффициент формулы для элемента с различными валентностями	-	-	Fe_p2_pfu, Fe_p3_pfu

Проверка на соответствие этой конвенцией и переименование полей выполняется автоматически при импорте и экспорте данных (см. раздел 2.3.1). Также было принято соглашение, в соответствии с которым, для именованя столбцов, содержащих соотношения элементов, используется префикс X_, например X_Fe_Mg_pfu.

2.3.5. Управление методами

Управление методами в рамках интерактивного реестра состоит в создании и редактировании страниц в соответствующих пространствах имен вики-системы. Для этого используются инструменты, встроенные в систему DokuWiki (рисунок 2.6). Требования к оформлению страниц методов рассмотрены в разделе 2.2.3.

Пустые шаблоны для оформления страниц методов размещены на страницах в корневом пространстве имен вики.

Для авторизации пользователей и управления правами на редактирование страниц также используются инструменты DokuWiki, в частности, встроенный в систему плагин ACL. Его настройками пользователи разделены на три группы: анонимные, имеющие доступ только к справочной части реестра, обычные пользователи, имеющие доступ к выполнению расчетов и редактированию страниц реестра и администраторы, имеющие доступ к управлению DokuWiki. Для обеспечения всех компонентов реестра информацией о сессии пользователя разработанный плагин к системе

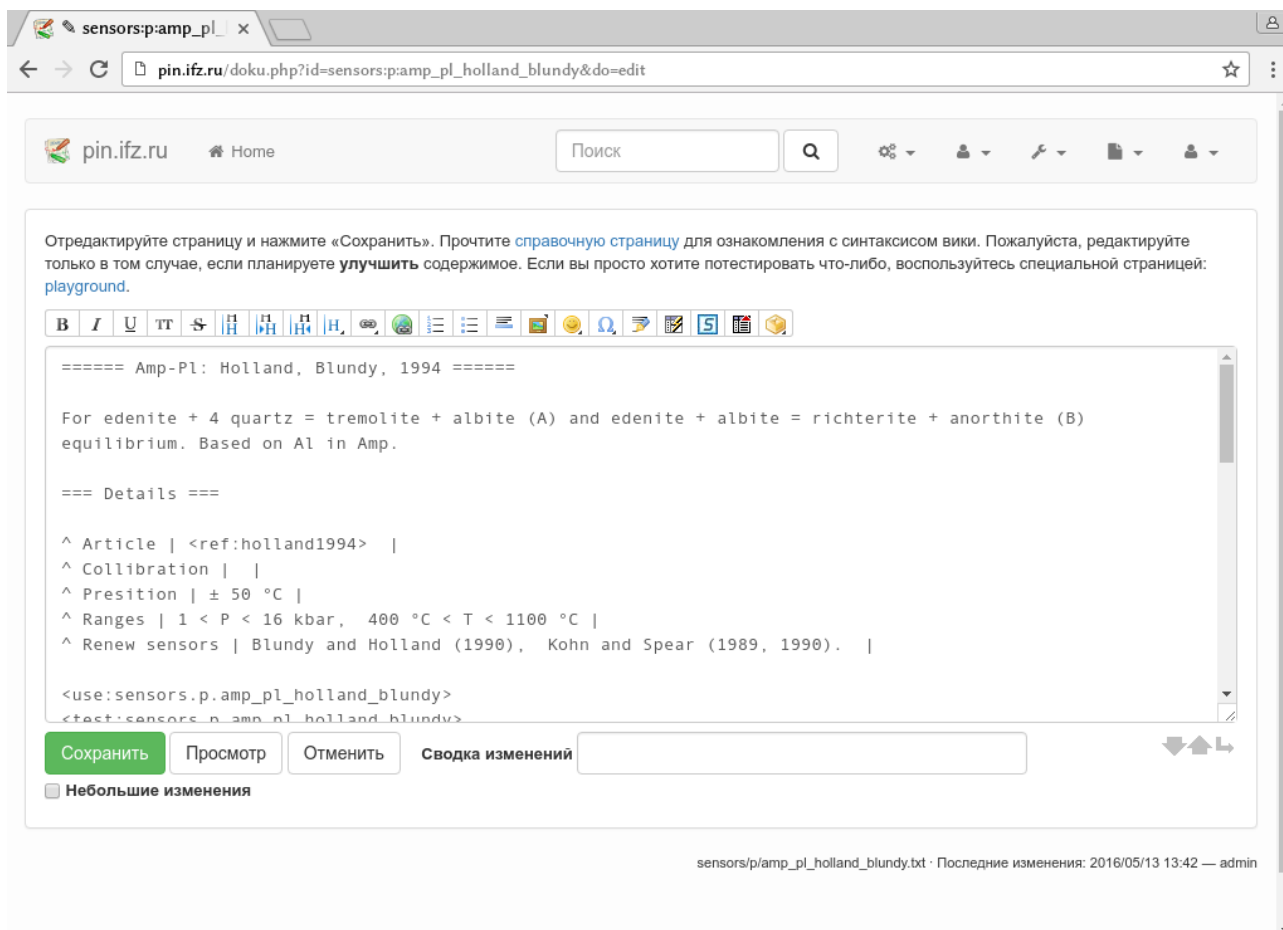


Рис. 2.6. Интерфейс редактирования страницы системы DokuWiki

DokuWiki передает эту информацию промежуточному серверу, который проверяет ее и принимает решение об обработке или отказе по тому или иному запросу пользователя.

2.3.6. Обеспечение безопасности при выполнении метода

Возможность пользователей самостоятельно добавлять методы в интерактивный реестр является потенциальной угрозой с точки зрения безопасности. В большинстве современных компьютерных систем выполнение произвольного стороннего кода рассматривается как недопустимая ситуация. Однако в данном случае оно необходимо для реализации концепции интерактивного реестра.

Для обеспечения безопасности при исполнении исходного кода методов были приняты следующие метры:

- Модульный сервер ни при каких обстоятельствах не исполняет пользовательский код. Всё исполнение кода производится на сервере Rserve;

- Rserve запущен в изолированном окружении с помощью механизма AppArmor [AppArmor Documentation, 2016];
- Для каждого запроса на расчет Rserve создает выделенный контекст исполнения, который уничтожается после его завершения (успешного или нет).

Изоляция пользовательского кода на уровне Rserve защищает от возможности использования промежуточного сервера злоумышленниками. Ограниченный профилем AppArmor в режиме принуждения (англ. enforce), сервер вычислений имеет доступ на чтение только к необходимым для его работы библиотекам. Он не может устанавливать сетевые соединения или получать доступ к файлами за пределами директории пользователя. Таким образом, достигается защита от компрометации операционной системы или модульного сервера. В самом Rserve существует механизм ограничения времени выполнения запросов, который защищает как от бесконечных циклов, возникших при добавлении в реестр недостаточно отлаженных алгоритмов, так и от реализации атак типа «отказ в обслуживании» (англ. DoS) через добавление вредоносного кода.

Следует отметить, что эти меры не отменяют, а дополняют механизм входа в систему, имеющийся в DokuWiki и используемый промежуточным сервером. Благодаря этому механизму вносить изменения в страницы и использовать методы расчета могут только зарегистрированные в вики-системе пользователи.

2.3.7. Пользовательский интерфейс выполнения расчетов

Интерфейс выполнения расчетов позволяет непосредственно применять имеющиеся в реестре методы для обработки пользовательских данных. Так как для различных категорий методов могут потребоваться различные элементы управления, то и пользовательский интерфейс может быть различен. Поэтому для каждой категории методов был создан свой модуль пользовательского интерфейса. Специально разработанный плагин обеспечивает интеграцию в указанные страницы вики-системы модуля пользовательского интерфейса написанного на JavaScript и использующего HTML и CSS для отображения данных. Через REST интерфейс он взаимодействует с соответствующим данной категории методов модулем промежуточного сервера (см. раздел 2.3). Таким образом, на одну категорию

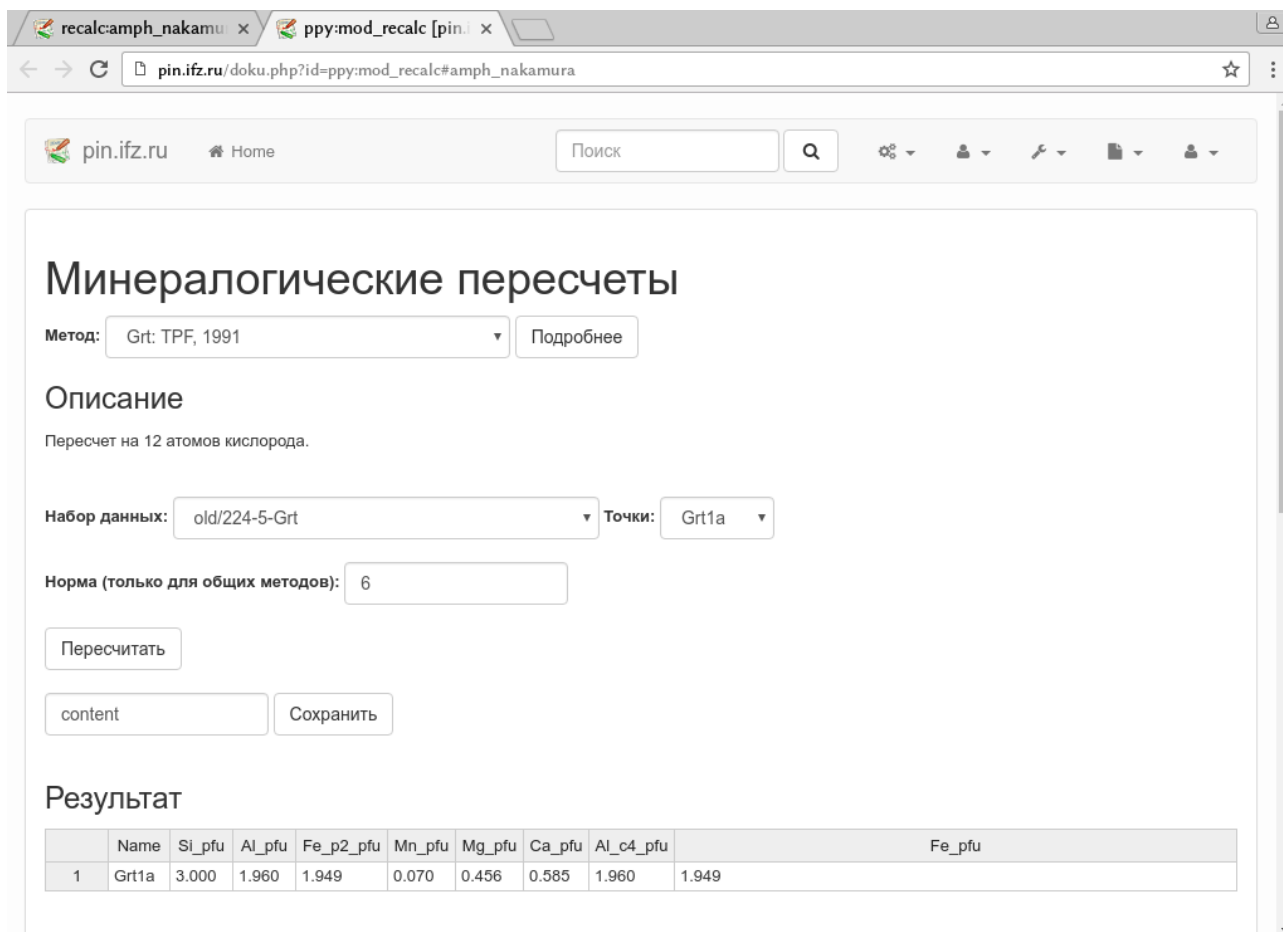


Рис. 2.7. Интерфейс модуля минералогических пересчетов

методов обработки приходится одно пространство имен в вики, один модуль промежуточного сервера и один модуль пользовательского интерфейса.

Модуль пользовательского интерфейса позволяет выбрать метод из имеющихся в реестре в рамках данной категории, выбрать исходные данные из загруженных пользователем, запустить расчет, а также отображает полученный результат. Интерфейс одного из таких модулей представлен на рисунке 2.7. Использование плагина к DokuWiki обеспечивает интеграцию интерфейса в страницу без использования механизма `iframe` или внесения изменений в код самой вики-системы. Он обслуживает модули пользовательского интерфейса всех категорий методов.

По умолчанию результаты расчета или графического построения не сохраняются, для их сохранения пользователь должен указать имя файла и нажать кнопку «Сохранить». После этого файл сохранится и будет доступен в папке пользователя под указанным именем.

Альтернативным подходом к построению пользовательского веб-интерфейса для приложений на R является Shiny [*Shiny — A web application*

framework for R, 2016] — библиотека для создания интерактивных веб-приложений с графическим пользовательским интерфейсом. Несмотря на ее привлекательность для использования в реестре, она требует применения R непосредственно для обработки HTTP запросов, что затрудняет интеграцию интерфейса на ее основе с вики-системой. Поэтому в данной реализации реестра она не используется.

2.3.8. Поддержка многоязычности

Поддержка многоязычности в интерактивном реестре является сложной проблемой. С одной стороны, для большинства современных приложений (в том числе веб-приложений) поддержка нескольких языков пользовательского интерфейса является нормой, с другой стороны вики-системы, как правило, одноязычны, т.к. сообщество, которое их развивает, обычно общается на одном языке (следует отметить, что даже Википедия не является исключением из этого правила, имея отдельные языковые разделы, связь между которыми обеспечивается специальными ссылками, известными как «интервики»).

В рамках работы над реестром было принято следующее решение: информация о методах в системе размещается на языке оригинала статьи или на английском, в то время, как язык интерфейса самой системы и модулей пользовательского интерфейса может переключаться.

DoKuWiki не имеет встроенного механизма поддержки многоязычности. Она добавляется с помощью специального плагина. Для реализации многоязычности на уровне модулей пользовательского интерфейса, в соответствии с существующими практиками, все локализуемые строки в модулях были вынесены в отдельные JavaScript файлы, загружаемые в зависимости от выбранного языка интерфейса. Выбор языка выполняется с помощью добавления пользователя в специальную группу. На данный момент для пользовательского интерфейса, кроме русского, поддерживается также английский язык.

2.4. Инструменты визуализации геохимических данных в интерактивном реестре

2.4.1. Базовые инструменты визуализации

Большинство рассмотренных в первой главе приложений для обработки петролого-геохимических данных имеют в своем арсенале ряд инструментов для их визуализации. В первую очередь, это базовые инструменты для построения диаграмм и гистограмм, а также тригонограмм или тройных диаграмм — распространенного в петрологии метода визуализации [Скляров, 2001]. Пример тригонограммы представлен на рисунке 2.8.

Хотя интерактивный реестр не является приложением для визуализации данных, наличие такой возможности является полезным для пользователя. Поэтому инструмент построения наиболее распространенных диаграмм включен в приложение в виде модуля (рис. 2.9).

Для анализа многопараметрических данных, в частности, о составе горных пород и минералов, была добавлена возможность построения диаграмм Эндрюса [Khattree, Naik, 2002]. При этом способе визуализации каждая строка данных представляется в виде кривой, причем строки с близкими значениями стягиваются в пучки. Данный метод позволяет быстро оценить наличие нескольких групп внутри имеющегося набора данных. Пример такой диаграммы представлен на рис. 2.10.

Помимо положения точек на координатных осях, можно использовать дополнительные средства визуализации — цвет и форму значков. Полученные графики сохраняются в векторном формате SVG, также поддерживается экспорт в другие форматы.

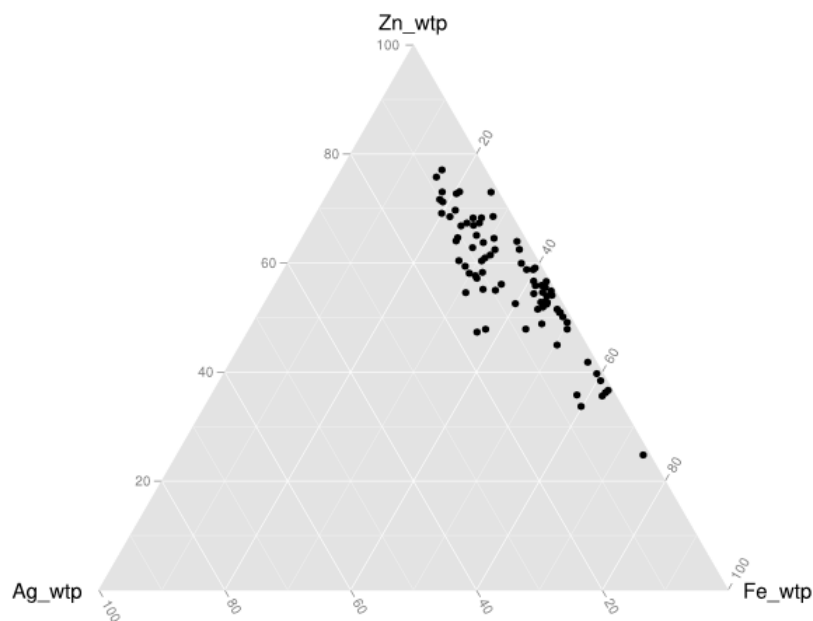


Рис. 2.8. Пример тригонограммы железо-цинк-серебро. Содержание дано в весовых процентах

Рис. 2.9. Интерфейс построения диаграмм. Показаны элементы управления для выбора типа графика (гистограмма, диаграмма, тригонограмма), столбцов данных и диапазонов

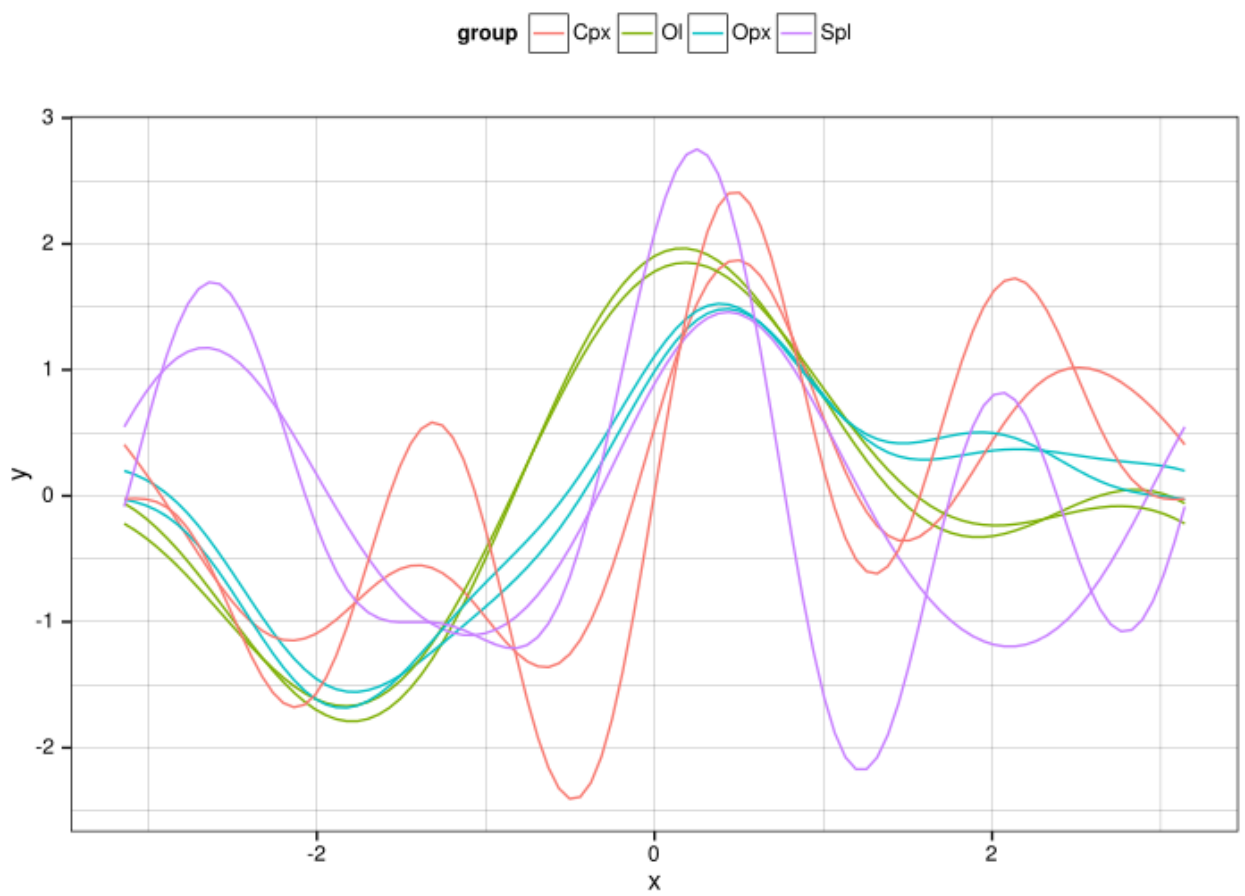


Рис. 2.10. Пример диаграммы Эндрюса [Khattree, Naik, 2002] для составов различных минералов

2.4.2. Специализированные инструменты визуализации петролого-геохимических данных

Кроме базовых инструментов визуализации, рассмотренных в предыдущем разделе, существуют специализированные, используемые при анализе петролого-геохимических данных. К ним относятся, в частности, дискриминационные диаграммы и спайдерграммы.

Дискриминационные диаграммы представлены диаграммами и тригонограммами, на которых заданы определенные области. Попадание результатов анализа в одну из таких областей отражает некоторые качества минерала или горной породы. Прежде всего, эти диаграммы используются для классификации (что и отражено в их названии). В качестве координат в дискриминационных диаграммах могут использоваться компоненты состава или их производные величины, например, суммы отдельных компонентов (рис. 2.11). Разнообразие таких диаграмм велико, при этом каждая дискриминационная диаграмма обладает своей областью применения [Скляр, 2001].

Другим распространенным методом, применяемым при визуальном анализе геохимических данных, являются спайдерграммы — диаграммы, полученные нормированием состава образца на определенный заданный состав, например, на состав примитивной мантии [McDonough, Sun, 1995]. При этом особый интерес представляет распределение редкоземельных элементов [Скублов, 2005] и, в частности, наличие аномалий [Pride, Muecke, 1981; Strekopytov, 1997]. Спайдерграммы выполняются в виде точек в прямоугольной системе координат, где по оси X задаются элементы в определенном порядке, а по оси Y — их относительное содержание в логарифмическом масштабе (рис. 2.12). Элементы по оси X перечисляются в определенном порядке, а точки соединяются линиями для каждого образца. Следует отметить, что в различных источниках этот порядок может быть различным.

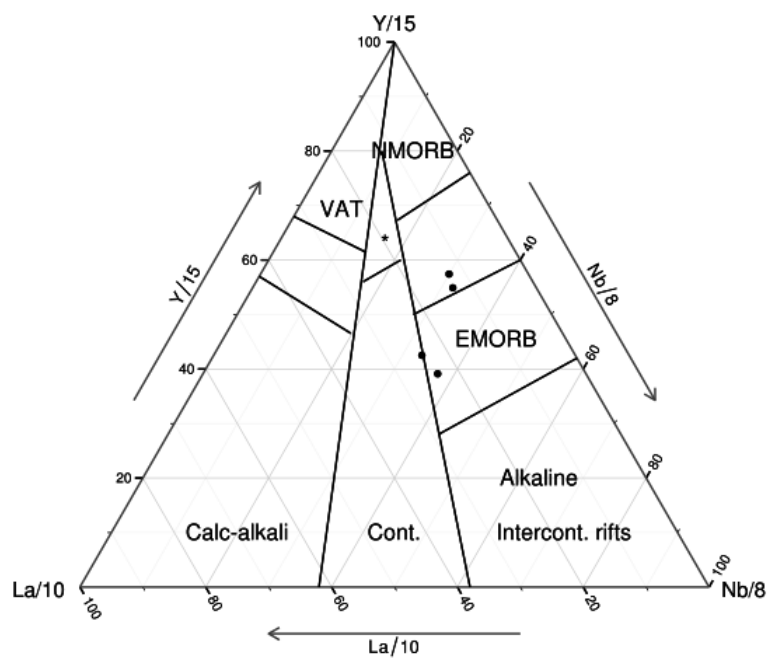


Рис. 2.11. Пример дискриминационной диаграммы для классификации по типам земной коры по [Cabanis, 1989]

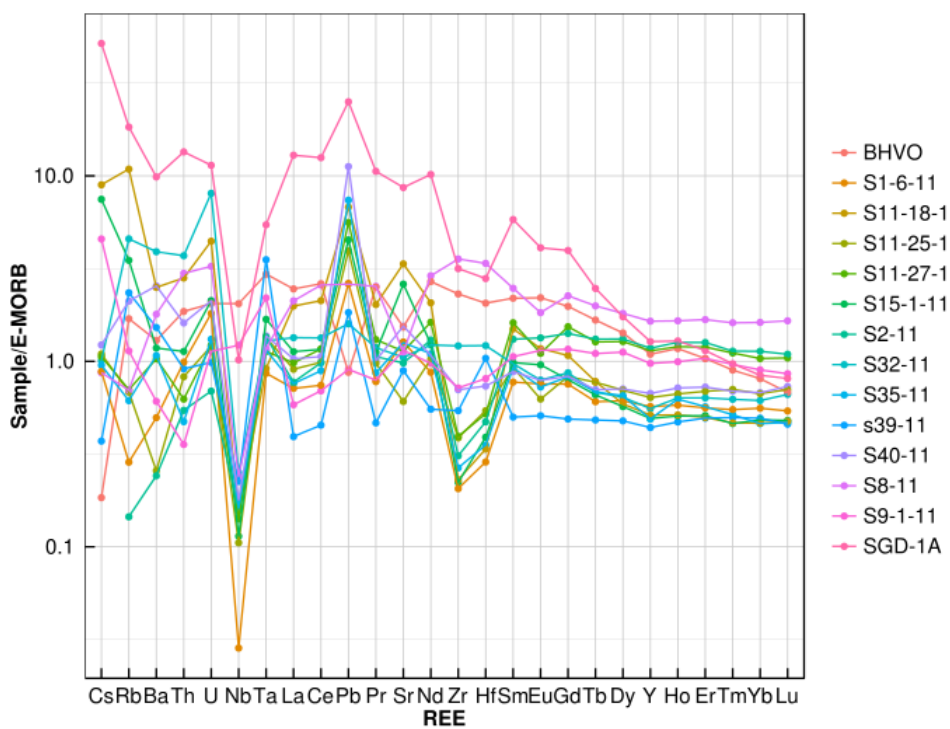


Рис. 2.12. Пример спайдерграммы. Нормирование содержания редкоземельных элементов на состав E-MORB по [Sun, McDonough, 1989]

2.4.3. Интеграция инструментов визуализации с интерактивным реестром

Базовые инструменты визуализации являются универсальными, поэтому для их включения в реестр не потребовалось создания отдельного пространства имен, содержащего описание методов построения. Были разработаны только модули пользовательского интерфейса и сервера.

Ввиду разнообразия специальных методов, в соответствии с концепцией интерактивного реестра, для хранения информации о них было выделено два пространства имен (для методов построения спайдерграмм и классификационных диаграмм), а также две пары модулей, обеспечивающие возможность их использования. Для построения графиков все модули используют библиотеку R `ggplot2` [Wickham, 2009; Wilkinson, 2006].

Взаимодействие модулей пользовательского интерфейса и сервера при построении всех типов диаграмм строится по общему принципу: после выбора типа построения, набора данных и настроек модуль отправляет на сервер POST запрос, содержащий опции указанные пользователем в формате JSON. Сервер выполняет построение и сохраняет изображение в директорию временных файлов. Модуль пользовательского интерфейса загружает изображение и отображает его, а также предоставляет возможность сохранить файл в директорию пользователя, присвоив ему имя.

Методы построения дискриминационных диаграмм записываются на языке R в виде функций, следующих общим правилам именования (см. раздел 2.3.3). В качестве входных параметров функция принимает набор данных, а возвращают объект, содержащий поля, описывающие элементы будущего построения. Структура объекта представлена в таблице 2.4.

Необходимость использования промежуточной структуры данных вместо прямого вызова функций графической библиотеки `ggplot2` обусловлено стремлением избавить исходный код методов от дополнительных зависимостей.

При построении дискриминационной диаграммы в отдельный файл записываются координаты точек, рассчитанные выбранным методом. Это позволяет экспортировать их в сторонние приложения для дальнейшей обработки и визуализации. Необходимость задавать координаты разделительных линий и подписей в виде исходного кода можно рассматривать

Таблица 2.4. Структура объекта, возвращаемого методами построения дискриминационных диаграмм

Поле	Значение
<code>type</code>	Базовый тип построения: диаграмма или тригонограмма
<code>points</code>	Точки для построения, содержит списки Name, X, Y и Z
<code>limits</code>	Задаёт границы построения, в виде вектора содержащего последовательно минимальное и максимальное значения для каждой оси
<code>marks</code>	Текстовые отметки на диаграмме, содержит списки Label, X, Y и Z
<code>lines</code>	Разделительные линии на диаграмме, содержит списки X1, Y1, Z1, и X2, Y2, Z2, соответствующие координатам начальных и конечных точек
<code>labels</code>	Содержит список подписей к осям

одновременно как преимущество и недостаток. С одной стороны, полученная на выходе диаграмма является полностью векторной и компактной по размеру. С другой, необходимость задания всех координат точек и подписей затрудняет добавление новых методов. Существует распространённая практика построения диаграмм поверх растровых изображений, содержащих разделительные линии. Однако из-за того, что качеством полученной диаграммы невозможно управлять, было принято решение об отказе от этого подхода в пользу полностью векторного. Кроме того, растеризация полученного изображения всегда возможна, в то время как векторизация является куда более сложной задачей.

Методы для построения спайдерграмм имеют более простой интерфейс. Помимо данных для построения, в качестве входного параметра они принимают порядок, в котором необходимо выстроить данные. Этот аргумент является опциональным. Возвращаемое значение содержит всего три поля: `values`, содержащее набор данных, прошедших нормирование, `grades` — вектор отметок по оси X и `labels`, содержащее подписи к осям. Как и для дискриминационных диаграмм, полученные координаты точек записываются в отдельный файл и могут быть использованы в дальнейшем.

2.5. Работа с методами вне реестра

Одним из направлений критики научных веб-приложений является высокая зависимость пользователей от разработчиков. В случае прекращения поддержки веб-приложения и приостановки работы его сетевой инфраструктуры конечные пользователи сразу же лишаются всех возможностей по его использованию, чего не происходит при использовании настольных приложений. Кроме того, для использования веб-приложения необходим постоянный доступ в Интернет.

Для преодоления, этих недостатков в реестр был добавлен механизм выгрузки имеющихся методов в виде библиотеки модулей R. После загрузки такой библиотеки на компьютер пользователя, он может использовать все доступные в реестре методы локально, используя настольную среду R, без необходимости подключения к сети Интернет. Формирование библиотеки происходит динамически в момент выполнения запроса на загрузку, поэтому она всегда содержит актуальные версии всех методов, доступных в реестре.

Естественным развитием этого подхода является создание соответствующего модуля для системы пакетов R и его публикация в репозитории CRAN. Для этого необходимо снабдить все методы тестовыми данными, что является одним из необходимых требований данного репозитория.

Другим способом работы с методами вне реестра является их удаленное использование через публичный API. Этот метод не реализован на данный момент в реестре, но является перспективным направлением для его развития. Использование REST в качестве интерфейса модульного сервера и распространенного формата JSON для передачи данных создают для этого хороший задел.

2.6. Выводы

В данной главе рассмотрена концепция интерактивного реестра методов обработки петролого-геохимических данных. Показана возможность ее реализации на базе веб-приложения. Рассмотрены выбор базовой платформы, организация хранения текстовой, библиографической и алгоритмической информации в ее рамках, инфраструктура выполнения расчетов. Результатом выбора платформы для построения реестра стала система DokuWiki, а язык R был выбран для описания алгоритмов обработки данных.

Для реализации концепции интерактивного реестра были разработаны следующие компоненты системы: промежуточный модульный сервер, плагин для DokuWiki, модули сервера и модули пользовательского интерфейса, профиль безопасности AppArmor для сервера RServe, механизм переводов. Большую часть пользовательского интерфейса составляет интерфейс DokuWiki, в рамках которого пользователь знакомится с доступными методами и осуществляет управление ими. Промежуточный сервер берет на себя следующие функции, отсутствующие в вики-системе, в том числе:

- Управление сервером вычислений (RServe);
- Загрузка в него актуальных исходных кодов методов со страниц вики;
- Обработка запросов от пользовательского интерфейса на проведение расчетов и их выполнение на сервере вычислений;
- Преобразование форматов файлов;
- Проверка прав пользователей при выполнении расчетов, загрузка в RServe исходных данных и сохранение результатов;
- Выгрузка библиотеки методов в виде архива для локального использования.

Разработанная модульная архитектура промежуточного сервера позволяет легко дополнять систему новыми категориями методов.

Предложенный подход и разработанная платформа могут быть применены не только в задачах обработки петролого-геохимических данных, но и в других областях, для которых характерно большое число вычислительных методов обработки сравнительно однородных наборов данных.

Использование механизма плагинов базовой вики-системы Dokuwiki позволило обойтись без внесения изменений в ее код, что позволяет задействовать механизм автоматического обновления, важный с точки зрения информационной безопасности.

Предпочтение связки из нескольких языков программирования для реализации реестра (PHP, Python, R, HTML/JavaScript) одной платформе обусловлен выбором вики системы DokuWiki (раздел 2.2.2) и R (раздел 2.3.2) в качестве языка представления методов. Выбор языка для реализации промежуточного сервера был произвольным, поэтому автор остановился на распространенном в веб-программировании языке Python3 [Rhodes, Goerzen, 2010]. Использование такого большого числа

разнообразных инструментов обусловлено удобством разработки. Все они активно развиваются на сегодняшний день и могут быть установлены на большинстве современных дистрибутивов ОС GNU/Linux непосредственно из соответствующих пакетных менеджеров. Минимальные версии, необходимые для работы системы: Linux 3.0, PHP 5.3, Python 3.4, R 3.2, Apache 2.4. Возможно также использование любого другого HTTP сервера с поддержкой PHP и обратного проксирования запросов (Nginx, lighttpd). Возможности по развертыванию реестра на базе ОС Windows ограничены необходимостью внесения изменений в метод отслеживания страниц в вики-системе и использованием иных методов изоляции процесса RServe [Urbanek, 2003].

Необходимость регистрации для работы в системе может оттолкнуть потенциальных пользователей реестра, однако отказаться от нее нельзя сразу по нескольким причинам: наличие файлов пользователя, необходимость контроля за внесением изменений в вики и защита от вандализма.

Глава 3. Интерактивный реестр методов геобаротермометрии

В связи с рассмотренными в разделе 1.5 проблемами существующих систем геобаротермометрии, актуальной задачей является создание информационной системы, объединяющей в себе большой набор различных минеральных геосенсоров и справочную информацию, необходимую для обоснованного выбора того или иного сенсора. Данная задача может быть эффективно решена на базе интерактивного реестра методов обработки петролого-геохимических данных. Для этого была выполнена формализация понятия геосенсор [Иванов, 2016а], определена необходимая дополнительная информация о нем [Иванов, 2015b] и реализованы соответствующие модули.

3.1. Формальное описание геосенсора

Для реализации потенциала реестра по организации методов геобаротермометрии необходимо формализовать понятие геосенсор в виде функции. Такая формализация будет в дальнейшем использована при создании API, обеспечив единую интерфейс вызова для различных сенсоров.

Общее определение геосенсора можно дать в виде следующей функции:

$$V = f(C_1, C_2, \dots, C_n, a_1, a_2, \dots, a_p), \quad (3.1)$$

где V — определяемое сенсором значение (температура в случае термометра и давление в случае барометра), f — функция сенсора, C_1, \dots, C_n — составы минералов, a_1, \dots, a_p — свободные параметры. Число минералов n , вовлеченных в расчет, может колебаться от одного до 4–5. Следует отметить, что в это число входят только те минералы, химический состав которых непосредственно используется в расчете, а не все минералы парагенезиса, для которого разработан сенсор. Количество свободных параметров p может быть равным нулю. В этом случае определяемое значение зависит только от химического состава. Однако, как правило, p равно 1. Тогда в качестве a_1 выступает второй термодинамический параметр (для геобарометра — температура, для геотермометра — давление). Также

возможно привлечение и других внешних параметров, например, активности флюида [Иванов, 2016а].

Вычисление значения функции f можно представить в виде нескольких последовательных этапов. Первым является переход к коэффициентам формул (к.ф.), которые, как правило, и используются при построении геосенсора. Вторым этапом является расчет дополнительных величин, таких, как содержание отдельных миналов и характеристические коэффициенты (например, магниальность, щелочность и т.п.), используемые впоследствии при расчете искомого параметра. Последним этапом является, собственно, расчет, в котором используются состав минералов, дополнительные и свободные параметры. Подобное разделение расчета на три этапа удобно, так как первые два из них определяются только минералом и могут быть использованы без изменений в нескольких сенсорах [Иванов, 2016а].

С алгоритмической точки зрения расчет, как правило, выполняется за константное время, однако в случае использования термодинамических соотношений при расчете может решаться задача минимизации. Кроме того, геотермометр и геобарометр для одного типа пород могут объединяться в геотермобарометр [Аранович, 1991]. При этом вычисление давления и температуры происходит итерационно, начиная с некоторых начальных значений:

$$\begin{cases} P = f_P(C_1, C_2, \dots, C_n, T', a_2, a_3, \dots) \\ T = f_T(C_1, C_2, \dots, C_n, P', b_2, b_3, \dots) \end{cases}, \quad (3.2)$$

где f_P и f_T — функции барометра и термометра соответственно, а P' и T' — значения давления и температуры, полученные в ходе предыдущей итерации. Тем не менее, использование ограничения на число итераций позволяет обеспечить не более чем константное время расчета для всех типов сенсоров.

3.2. Реализация на базе интерактивного реестра

В соответствии с концепцией интерактивного реестра, для хранения информации о геосенсорах было выделено несколько пространств имен, по одному для термометров, барометров, термобарометров и фугометров. Для каждого метода в соответствующем пространстве имен создается отдельная страница. Кроме того, если определенные этапы расчета

являются одинаковыми для многих сенсоров, то их алгоритм может быть размещен в виде отдельного метода в другом пространстве имен. Таким образом, кроме пространства имен `:sensors:` и вложенных в него (`:sensors:p:`, `:sensors:t:`, `:sensors:pt:`, `:sensors:f:`) были задействованы также пространства имен `:recalc:` и `:interpmin:`. При этом для именованя страниц сенсоров используется принятое соглашение (см. раздел 2.2.3).

Выбор метода для выполнения конкретного геохимического расчета может быть сложной для пользователя задачей, поэтому страница сенсора в реестре должна содержать определенный набор справочной информации [Иванов, 2015b]:

- Информация о публикации: библиографическая ссылка на статью, по возможности — ссылка на полнотекстовую версию статьи (с использованием механизма управления библиографическими ссылками реестра, см. раздел 2.2.4);
- Информация о парагенезисе, для которого предназначен данный сенсор;
- Точность метода;
- Пределы входных и выходных значений;
- Информация о наборе данных, на которых выполнялась калибровка;
- Список геобаротермометров и/или их калибровок, которые могут рассматриваться как устаревшие в связи с публикацией данного сенсора;
- Особые указания по использованию сенсора, данные автором и, возможно, другими пользователями.

Шаблон страницы геосенсора соответствует общему шаблону показанному в листинге 2.1. Формальные данные, такие как точность и парагенезис, указываются в таблице (аналогично строке 8 листинга). Оформление исходного кода метода, записанного на языке R, соответствует шаблону 2.3, рассмотренному в разделе 2.3.3. Для каждого метода определяется три функции: непосредственно выполняющая расчет, возвращающая список используемых минералов и выполняющая тестирование правильности работы метода. Также в страницу включается ссылка, позволяющая сразу же перейти в интерфейс выполнения расчета для обработки

пользовательских данных с его помощью. Вместе с ней располагается ссылка для перехода к инструменту тестирования. Информация о состоянии страницы описания метода включает список задач, необходимых для ее совершенствования. Дополнительная информация о методе может включать комментарии, дополнительные ссылки и сведения об опыте использования данного инструмента. Эта часть страницы является наименее формальной. Пример страницы описания сенсора приведен на рисунке 3.1.

Для выполнения расчетов с использованием геосенсоров были разработаны модули пользовательского интерфейса и сервера. Они обслуживают все типы геосенсоров. Особенность работы модуля сервера состоит в том, что он взаимодействует с исходным кодом метода не только на этапе выполнения расчета. При выборе метода модуль получает список минералов, составы которых используются сенсором и список свободных параметров, через вызов соответствующих функций. После получения этих данных он отображает поля ввода в соответствии с полученными данными. Интерфейс модуля представлен на рисунке 3.2.

sensors:t:amp_pl_holland_blundy [pin.ifz.ru] – Chromium

pin.ifz.ru/doku.php?id=sensors:t:amp_pl_holland_blundy

Amp-Pl: Holland, Blundy, 1994

For edenite + 4 quartz = tremolite + albite (A) and edenite + albite = richterite + anorthite (B) equilibrium.

Details

Article	Tim Holland, Jon Blundy <i>Non-ideal interactions in calcic amphiboles and their bearing on amphibole-plagioclase thermometry</i> , 1994 [ref]
Collibration	
Precision	± 50 °C
Ranges	1 < P < 16 kbar, 400 °C < T < 1100 °C
Renew sensors	Blundy and Holland (1990), Kohn and Spear (1989, 1990).

Использовать Проверить

- заголовок
- краткое описание
- статья
- ссылка
- карточка
- подробное описание
- кед
- тест

Source

```

1. source('recalc_amp_holland_blundy.r');
2. source('recalc_pl_8ox.r');
3. source('interpmin_amp_holland_blundy.r');
4.
5. sensors.t.amp_pl_holland_blundy <- function(Amp, Pl, P, usePFU = FALSE, ...){
6.
7.   Amp1 <- recalc.amp_holland_blundy(Amp);
8.   Amp2 <- interpmin_amp_holland_blundy(Amp);

```

Рис. 3.1. Пример страницы геосенсора: амфибол-плагиоклазовый термометр [Holland, Blundy, 1994].

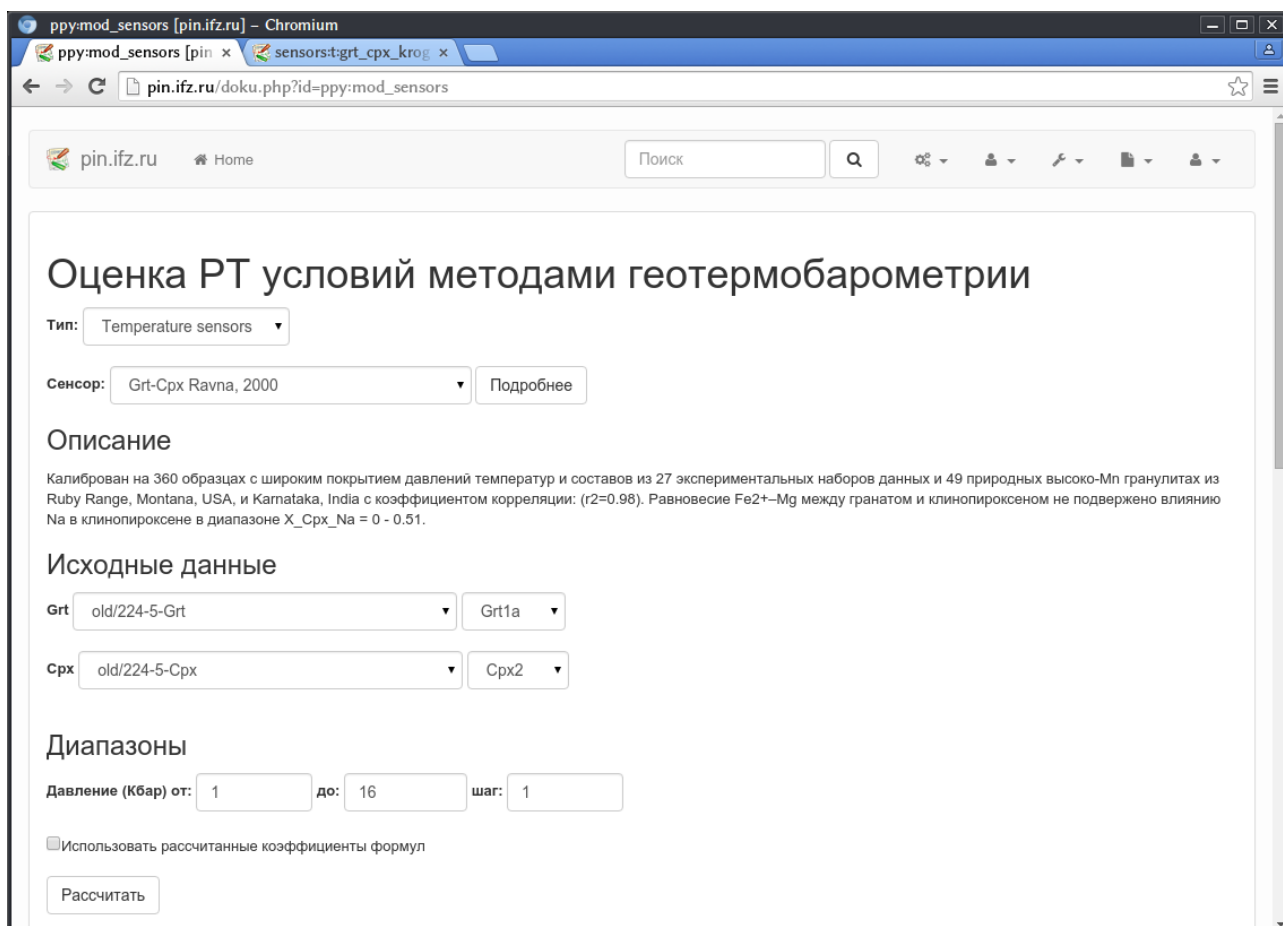


Рис. 3.2. Интерфейс модуля геотермобарометрии. Выбран гранат-клинопироксеновый сенсор [Ravna, 2000]. Показано краткое описание сенсора и поля для задания диапазона свободного параметра (давление)

3.3. Наполнение базы методами

Следующим после разработки соответствующих модулей интерактивного реестра этапом работы стало наполнение базы различными геосенсорами. В качестве источников алгоритмов сенсоров использовались оригинальные статьи, а также база программы TRF (см. раздел 3.3.1).

При ручном наполнении базы автор столкнулся с рядом сложностей. Во-первых, во многих оригинальных статьях нет контрольных примеров входных и выходных данных, что ограничивает использование техники разработки через тестирование (Test-driven Development, TDD) [Beck, 2003] при добавлении методов в реестр. Другой проблемой является частое отсутствие в статьях прямых указаний на методы пересчета, применяемые к минералам. Эти две проблемы затрудняют добавление описанного в статье метода в реестр. С другой стороны, в качестве информационного приложения к многим современным статьям, посвященным геосенсорам, идут электронные таблицы, содержащие алгоритм расчета (например, [MacGregor, 2015b] или [Nakamura, 2015]). Наличие таких таблиц позволяет добавлять метод сравнительно быстро.

Еще одной проблемой является недостаточное описание требований и ограничений для данного сенсора в оригинальной статье. Такая информация, как допустимые содержания элементов или максимальные и минимальные значения оцениваемой величины, может быть представлена в статье в неявном виде (например, в виде границ устойчивости данной минеральной ассоциации), что требует внимательного анализа и, возможно, консультации со специалистом.

3.3.1. Импорт базы сенсоров программы TRF

Для расширения списка поддерживаемых сенсоров была использована база инструментов программы TRF [Fonarev, Graphchikov, Konilov, 1991] с разрешения одного из авторов А.Н. Конилова. Она включает более 200 различных инструментов.

В программе TRF для описания сенсоров используется специальный формальный язык. В качестве примера рассмотрим описание сапфирин-шпинелевого термометра Оуэна и Грино [Owen, Greenough, 1991] на этом языке (листинг 3.1). Описание состоит из нескольких секций. В

секции Reference задается название сенсора (фамилии авторов и год), в секции Mineral assemblage — минеральная ассоциация для данного сенсора, в секции Mole ratio задается метод пересчета и объявляются дополнительные переменные. Например, в строке 4 значение S задается равным содержанию кремния в сапфирине при пересчете по умолчанию (о чем свидетельствует команда <SPR>), умноженному на два. Секция Equations содержит непосредственно формулы для получения оценки температуры — значения переменной T. В случае наличия независимых параметров, они размещаются в специальной секции Input parameter (в данном листинге эта секция отсутствует).

```

1 Reference: OWEN J.V., GREENOUGH J.D. (1991)
2 Mineral assemblage: SPR+SPL
3 Mole ratio:      <SPR> A:=(AL4+AL6)*2;
4                  <SPR> S:=SI*2;
5                  <SPR> CT:=(CR+2*TI)*2;
6                  <SPR> F:=(FE2+FE3)*2;
7                  <SPR> M:=MG*2;
8                  <SPL> Y:=FE2/MG;
9 Equations:
10  A4:=6-S;
11  A6:=A-A4;
12  if A4>(A6+CT);
13    F3:=A4-A6-CT;
14    F2:=F-F3;
15  end;
16  if A4<(A6+CT);
17    F2:=F;
18  end;
19  KD:=Y/(F2/M);
20  T:=800+228*LN(KD)-273.15;

```

Лист. 3.1. Пример описания сенсора в программе TRF

Импорт сенсоров из библиотеки программы TRF был выполнен в два этапа. Вначале, с использованием инструмента PyParsing [McGuire, 2008], был разработан синтаксический анализатор, выполняющий разбор файлов описания сенсоров программы TRF. Файл, содержащий описания инструментов, разбивается на отдельные описания, далее каждое из них делится на секции (Reference, Mineral assemblage и т.д.). Содержание каждой секции разбирается в соответствии со своим синтаксисом. На втором этапе был разработан набор функций обратного вызова для

данного анализатора. Эти функции генерируют компоненты страницы интерактивного реестра. Получаемая на выходе страница содержит все необходимые структурные элементы и подготовленные поля для вставки дополнительных данных о сенсоре при этом также выполняется трансляция исходного кода сенсора на язык R.

Пересчеты в программе TRF реализованы в виде внутренних функций, но, благодаря доступу к исходным кодам, любезно предоставленным А. Н. Кониловым, эти функции были вручную переведены на R и добавлены в интерактивный реестр в виде модуля `utilities_tpf.r`. Вызов этих функций включается в генерируемый исходный код.

3.4. Использование реестра на примере фенгитовой мономинеральной барометрии

Для глубокометаморфизованных комплексов часто фиксируется неоднократное проявление метаморфических процессов, что затрудняет выделение равновесных минеральных ассоциаций в породе [Камзолкин, Иванов, Конилов, 2015]. В частности, эта проблема существует для гнейсов блыбского метаморфического комплекса Передового хребта Большого Кавказа, что не позволяет использовать распространенные геобарометры, например, Grt-Opx-Pl-Qtz [Perkins, Chipera, 1985] или Grt-Cpx-Pl-Qtz [Eckert, Newton, Kleppa, 1991].

Представленные в литературе оценки пикового давления метаморфизма комплекса в целом (8-9 Кбар по [Петрология ..., под. ред. Кориковского, 1991]) отличаются от оценок для эклогитов данного комплекса (до 18 Кбар по [Perchuk, Philippot, 1997]). Изучение включений омфациита в сульфидах гнейсов блыбского комплекса является аргументом в пользу его высокобарного происхождения [Камзолкин, 2013]. Другим аргументом могут служить результаты мономинеральной фенгитовой геобарометрии полученные в ходе работы [Камзолкин, Иванов, Конилов, 2015], в которой использовался прототип интерактивного реестра.

Фенгит $K(Al, Mg)_2(Si, Al)_4O_{10}(OH)_2$ — распространенный минерал, устойчивый в широком диапазоне давлений. Он относится к группе белых слюд и является твердым раствором ряда мусковит-селадонит по классификации, предложенной в работе [Tischendorf et al., 2007]. Изоморфное замещение кремния алюминием в этом твердом растворе

подчиняется обмену Чермака: $(\text{Mg}, \text{Fe}^{2+}), \text{Si} \leftrightarrow \text{Al}^{\text{IV}}, \text{Al}^{\text{VI}}$. Также в фенгите возможно ограниченное замещение калия натрием [Miyashiro, Shido, 1985].

Экспериментальные исследования зависимости составов фенгита от P-T-условий проводились для системы KFMASH ($\text{K}_2\text{O}-\text{MgO}-\text{Al}_2\text{O}_3-\text{SiO}_2-\text{H}_2\text{O}$) [Massonne, Schreyer, 1987], KFASH ($\text{K}_2\text{O}-\text{FeO}-\text{Al}_2\text{O}_3-\text{SiO}_2-\text{H}_2\text{O}$) [Massonne, Szpurka, 1997]. Кроме того, на основе модели фенгита [Coggon, Holland, 2002] и экспериментальных данных по фенгитам была получена формула фенгитового барометра для давлений, не превышающих 12 Кбар [Caddick, Thompson, 2008].

В рамках работы [Камзолкин, Иванов, Конилов, 2015] была проведена проверка возможности использования фенгитовой геобарометрии для более широких диапазонов давлений. Было проведено тестирование существующих сенсоров на большом объеме литературных данных. В прототип интерактивного реестра были добавлены фенгитовые барометры Каддика — Томпсона [Caddick, Thompson, 2008] и Массона — Шрайера [Massonne, Schreyer, 1987], а также были подготовлены исходные данные по 491 природному образцу из обобщающей около 150 публикаций работы Чамберса и Кона [Chambers, Kohn, 2012] и по 156 экспериментальным образцам из различных источников. Было выполнено обзорное построение исходных данных в осях Si_{rfu} , T и P. На полученном графике (рис. 3.3) можно выделить три группы образцов, отличающихся характером зависимости давления от остальных параметров.

После этого данные были сгруппированы и использованы для получения оценок давления с помощью упомянутых выше барометров. Поскольку исходные данные содержали также значение давления для всех образцов, то имелась возможность сравнить их с полученными результатами. Оценки, полученные с использованием сенсора Массона — Шрайера, являются слабо заниженными для первой группы и существенно заниженными для второй группы (до 10 Кбар). Сенсор Каддика — Томпсона дает оценки давления, завышенные в среднем на 2,5 Кбар. Для третьей группы, при > 35 Кбар ошибка для всех методов велика (> 7 Кбар) [Камзолкин, Иванов, Конилов, 2015].

Далее методом линейной регрессии были выведены уравнения, характеризующие зависимость давления от температуры и состава для каждой из групп ([Камзолкин, Иванов, Конилов, 2015]). Они также были

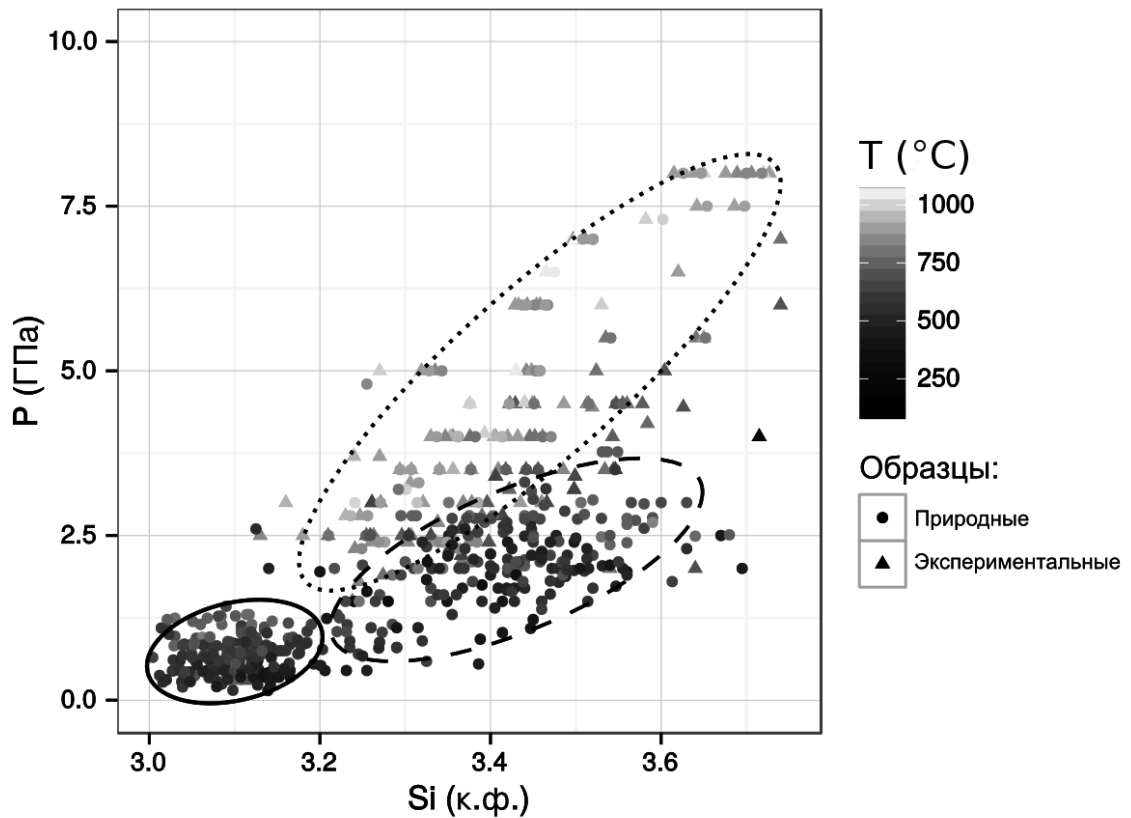


Рис. 3.3. Три группы образцов фенгита в зависимости от содержания кремния и температуры: 1. $Si < 3.2$ к.ф., $T < 750$ °C; 2. $Si > 3.2$ к.ф., $T < 750$ °C; 3. $T > 750$ °C по [Камзолкин, Иванов, Конилов, 2015]

добавлены в реестр. Сравнение оценок, полученных с помощью сенсоров и предложенных уравнений, представлены на рисунке 3.4.

Затем полученные формулы были применены к данным по 48 составам фенгита из пород, характеризующих основные разрезы блыбского комплекса. Расчет давлений проводился для температуры $T = 680$ °C, установленной для пиковых условий проградного метаморфизма эклогитов урочища Красной Скалы на р. Уруштен [Perchuk, Philippot, 1997]. С учетом погрешности минимальное давление, рассчитанное для второй группы фенгитов, составляет 15 Кбар, что примерно соответствует оценкам давления, полученным для эклогитов Красной Скалы [Perchuk, Philippot, 1997]. Этот факт может служить дополнительным доводом в пользу внутренней когерентности блыбского комплекса и предполагает совместный для эклогитов и вмещающих пород метаморфизм на уровне эклогитовой фации [Камзолкин, Иванов, Конилов, 2015].

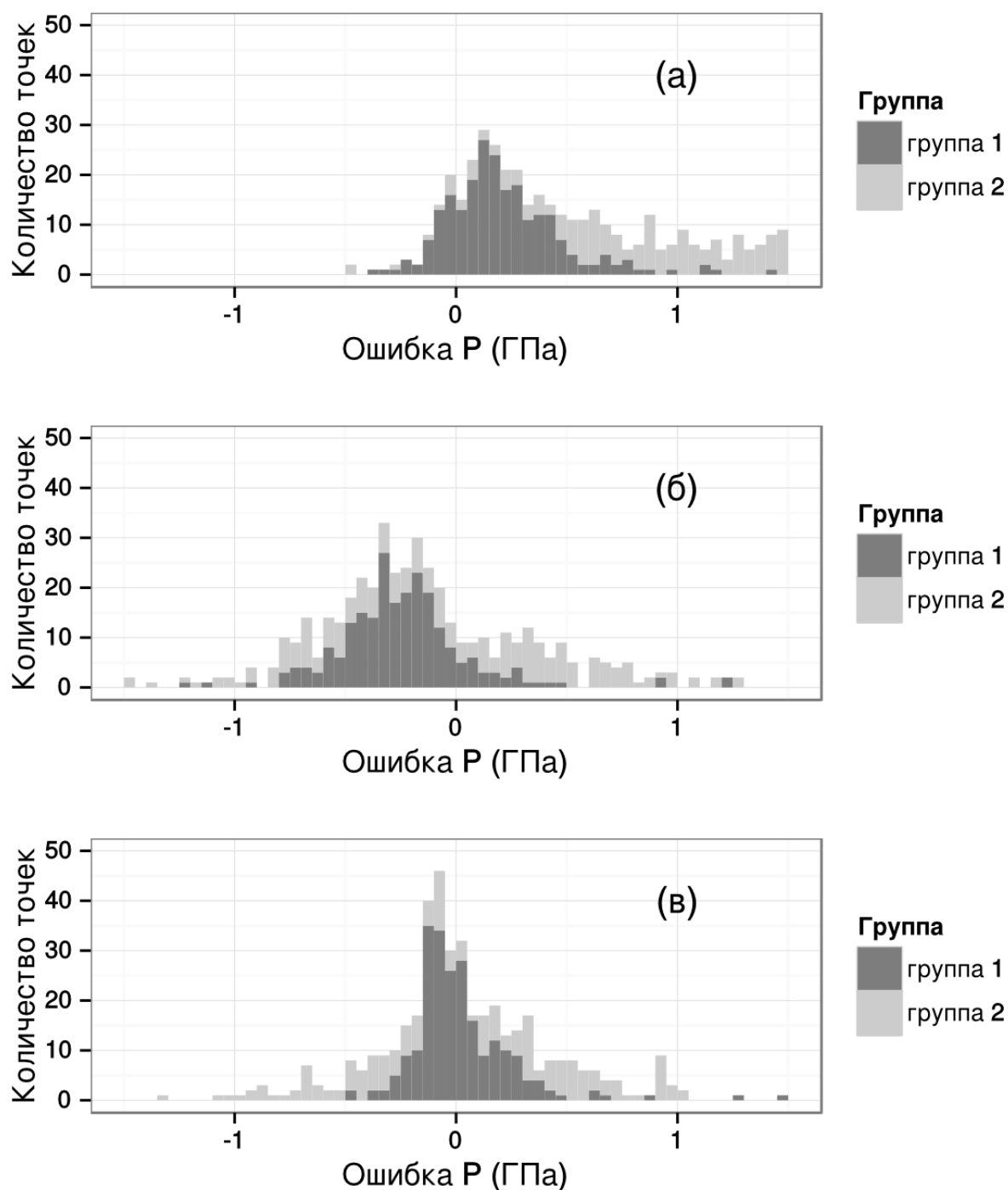


Рис. 3.4. Сравнение абсолютной ошибки определения давления ($P - P_{calc}$) с использованием фенгитовых мономинеральных барометров: а — барометр Массона — Шрайера [Massonne, Schreyer, 1987], б — барометр Каддика — Томпсона [Caddick, Thompson, 2008], в — предложенные в работе [Камзолкин, Иванов, Конилов, 2015] сенсоры для первых двух групп образцов (см. рис. 3.3)

3.5. Выводы

Формализация геосенсора в виде набора функций от определенного набора аргументов делает возможным добавление методов геобаротермометрии в интерактивный реестр, для чего был разработан соответствующие модули.

Использование вики-системы в качестве платформы для реестра открывает возможности его использования в качестве технической базы для сравнения и обсуждения различных методов обработки петролого-геохимических данных. В частности, интерес представляет определение наиболее актуальных инструментов, основанное как на анализе ссылочной информации, так и на непосредственном сравнении точности для различных наборов геохимических данных.

Другим направлением является взаимная верификация методов и составление наборов методов, дополняющих друг друга в рамках реестра, без необходимости создания отдельных приложений для их использования. Такие наборы могут быть организованы в соответствии с принципом внутренней согласованности, как предложено в работах [Fonarev, Graphchikov, Konilov, 1991] и [Spear, 1995] или по применимости для определенных типов горных пород.

Наполнение реестра связано с рядом сложностей, в частности, не все статьи, посвященные тем или иным методам геобаротермометрии, содержат весь объем информации, необходимой для воспроизведения метода в виде исходного кода. Тем не менее, в ходе работы с дополнительными литературными источниками возможно как точное воссоздание алгоритма метода, так и получение необходимых тестовых данных.

Прототип реестра был использован в работе [Камзолкин, Иванов, Конилов, 2015] для сравнения нескольких мономинеральных фенгитовых барометров и получения новых оценок давления для вмещающих гнейсов блыбского метаморфического комплекса Передового хребта Большого Кавказа.

Глава 4. Система идентификации минералов по результатам химического анализа

Помимо методов геобаротермометрии и вспомогательных инструментов, на базе интерактивного реестра была реализован еще один инструмент для обработки геохимических данных. Он предназначен для идентификации минералов по результатам химического анализа.

К информации о минеральном составе образцов предъявляются высокие требования, в частности, в задачах геобаротермометрии [Перчук, Рябчиков, 1976]. Кроме того, некоторые геосенсоры накладывают дополнительные ограничения на химический состав минералов, поэтому перед их применением может потребоваться дополнительный отбор анализов. Визуальное определение минералов в микрошлифе горной породы может быть затруднено [Reed, 2005]. Поэтому полезной является возможность идентификации и контроля минерального состава образцов после проведения микросондового анализа. Следует отметить, что в рамках идентификации минерала по химическому составу невозможно различить полиморфные модификации минералов. Кроме того, использование такой системы для определения произвольных минералов может быть затруднено в связи с отсутствием или недостаточным количеством наборов ограничений (см. раздел 4.3). Поэтому основными задачами, на решение которых ориентирована данная система, являются отбор образцов по химическому составу в соответствии с заданными требованиями и выявление недопустимости имеющегося состава для предполагаемого минерала.

4.1. Идентификация минералов средствами нечеткой логики

Под идентификацией минерала в данной работе подразумевается установление соответствия между химическим составом образца и его номенклатурным названием. Для этого необходимо каким-либо образом произвести сравнение химического состава данного образца с составом, известным заранее. Например, можно ввести ограничения на относительное

содержание тех или иных элементов в составе минерала. При этом за основу при составлении таких ограничений может быть взята теоретическая или эмпирическая формула минерала. А для оценки степени соответствия можно использовать аппарат нечеткой логики [Заде, 1976].

При использовании формулы минерала как основы для составления ограничений удобно перейти к расчетам в коэффициентах формул (к.ф.) — количеству атомов того или иного химического элемента на структурную единицу минерала. Наиболее простой способ перехода к к.ф. состоит в выборе элемента или элементов, содержание которых является постоянным для структурной единицы данного минерала, и нормировании состава на них:

$$C_{pfi} = \{c_i \cdot N / (c_k + \dots + c_m)\}, \quad (4.1)$$

где C_{pfi} — состав образца в к.ф.; N — количество атомов, на которое производится нормирование; c_i — содержание i -го элемента в атомных процентах (число атомов элемента в единице вещества в процентах); а $c_k + \dots + c_m$ и т.д. — содержания элементов (k, \dots, m -го), на сумму которых производится нормирование [Иванов, 2016с].

Для нормирования силикатных минералов обычно используется общее количество атомов кислорода, а также сумма катионов или анионов в кристаллохимической формуле минерала. Эти способы пересчета являются примитивными по сравнению с более сложными специализированными алгоритмами для определенных групп минералов. Например, существенно более сложный алгоритм пересчета для амфиболов использован в работе [Holland, Blundy, 1994]. Однако, несмотря на свою примитивность, они являются универсальными. Таким образом, для каждого минерала набор ограничений должен включать тип пересчета, число атомов N , на которое он производится, и правила, ограничивающие его состав.

Рассмотрим составление набора ограничений для биотита $K(\text{Mg, Fe, Mn, Al, Ti})_3\text{Al}(\text{Al, Si})\text{Si}_2\text{O}_{11}(\text{H}_2\text{O, F})$. Для данного минерала класса силикатов будем использовать метод пересчета на кислород, при $N = 12$. При этом мы получим следующие ограничения:

- $K = 1$;
- $3 \leq \text{Mg, Fe, Mn, Al, Ti} \leq 7$;
- $1 \leq \text{Al} \leq 5$;

- $0 \leq \text{Fe, Mg, Mn, Ti} \leq 3$;
- $2 \leq \text{Si} \leq 3$.

Кроме этих ограничений необходимо ввести дополнительное ограничение на содержание примесей в составе биотита. Оно может быть сформулировано следующим образом: «содержание не учтенных при составлении ограничений элементов не должно превышать определенного предела».

Еще один тип ограничений необходим для разделения минералов, наименование которых зависит от преобладающих элементов в группах замещения. Например, для минеральной группы граната такие ограничения обеспечат выбор между альмандином, пиропом, гроссуляром и т.д. В листинге 4.1 приведены полученные для биотита ограничения на формальном языке. Первые четыре строки носят информационный характер, в строке 5 представлена информация о нормировании, а остальные строки содержат ограничения. Такой способ записи понятен человеку и может быть непосредственно использован программой для построения и использования правил.

```

1 mineral: Bt
2 name: Biotite
3 id: Dana,71.02.02b
4 formula: K(Mg,Fe,Mn,Al,Ti)3 Al (AlSi) OSi211 + (H2O,F)
5 norm: 0,12
6 rule: K=1
7 rule: Mg,Fe,Mn,Al,Ti = 3~7
8 rule: Al = 1~5
9 rule: Fe,Mg,Mn,Ti = 0~3
10 rule: Si = 2~3
11 elements: O,K,Mg,Mn,Fe,Ti,Al,Si

```

Лист. 4.1. Ограничения, налагаемые на состав биотита

Чтобы воспользоваться аппаратом нечеткой логики, необходимо перейти от ограничений к функциям принадлежности. Такие функции были

заданы для всех трех групп ограничений. Функция принадлежности для ограничений на содержание элементов в структурной формуле задается следующим образом:

$$F(V, A, B) = \begin{cases} 0, & V < A - \beta \\ 0, & V - A < 0 \\ 1 - (A - V)/\beta, & A - \beta \leq V < A \\ 1, & A \leq V \leq B \\ 1 - (V - B)/\beta, & B < V \leq B + \beta \\ 0, & V > B + \beta \end{cases}, \quad (4.2)$$

где V — результат вычислений в левой части ограничения; A и B — верхняя и нижняя граница допустимого содержания элемента в к.ф. Данная функция принадлежности имеет трапециевидный график (который в случае $A = B$ вырождается в треугольник), причем угол наклона боковых сторон трапеции зависит от значения параметра β . Функция принадлежности, порождаемая ограничением на содержание примесей, задается следующим образом:

$$F(W) = \begin{cases} 1 - (\Sigma w_i/N), & \Sigma w_i/N < 1 \\ 0, & \Sigma w_i/N \geq 1 \end{cases}, \quad (4.3)$$

где Σw_i — сумма содержаний всех элементов, кроме перечисленных в соответствующем ограничении в к.ф.; N — число атомов, на которое производилось нормирование при переходе к к.ф. Деление на N позволяет устранить проблему, связанную с необоснованным уменьшением значения $F(W)$ при больших значениях N , например, при нормировании на 24 атома кислорода. Функция принадлежности для ограничений на порядок в группах замещения задается как:

$$F(V_1, V_2) = \begin{cases} 1, & V_1 \geq V_2 \\ \tau, & V_1 < V_2 \end{cases}, \quad (4.4)$$

где V_1 и V_2 — результаты вычисления выражений в правой и левой части ограничения соответственно; τ — коэффициент, определяющий, насколько сильно правила из этой группы должны влиять на конечный результат [Иванов, 2016с].

Для оценки степени соответствия между набором ограничений, описывающих минерал, и данным химическим составом вводится результирующая функция в виде логического умножения всех функций принадлежности, полученных из набора ограничений для данного минерала:

$$R = F_1 \wedge \dots \wedge F_n, \quad (4.5)$$

где R — результирующая функция; $F_1 \wedge \dots \wedge F_n$ — функции принадлежности, порожденные соответствующими ограничениями. Операция логического умножения выполняется в соответствии с правилами нечеткой логики. Значение результирующей функции интерпретируется как степень соответствия химического состава минерала данному набору ограничений. Для этого необходимо задать пороговое значение R , при котором на вопрос о соответствии все еще дается положительный ответ [Иванов, 2016с].

4.1.1. Определение параметров

Для определения параметров и порогового значения результирующей функции были использованы данные микронзондовых измерений, любезно предоставленные А. Л. Кулаковским (ИФЗ РАН) и содержащие результаты 1428 анализов для 119 различных образцов. Для всех анализов минеральный состав был заранее определен. Анализ составов минералов был выполнен в Лаборатории локальных методов исследования вещества МГУ.

Для обоснованного выбора порогового значения результирующей функции и выбора параметров β и τ были выбраны наиболее часто встречающиеся в данном наборе геохимических данных минералы — гранат (549 анализов), плагиоклаз (371 анализ) и биотит (356 анализов). Для этих минералов на представленном выше формальном языке были составлены наборы ограничений. После этого программа, интерпретирующая эти ограничения и вычисляющая значение результирующей функции, была применена к анализам этих минералов. Было выполнено несколько прогонов с различной величиной порогового значения R от 0.5 до 1 и различными значениями параметра β . При этом оценивалась относительная доля ложноположительных и ложноотрицательных результатов. Результат подбора представлен на рисунке 4.1. Исходя из этого графика, пороговое значение R было принято равным 0.88. Параметр β был установлен

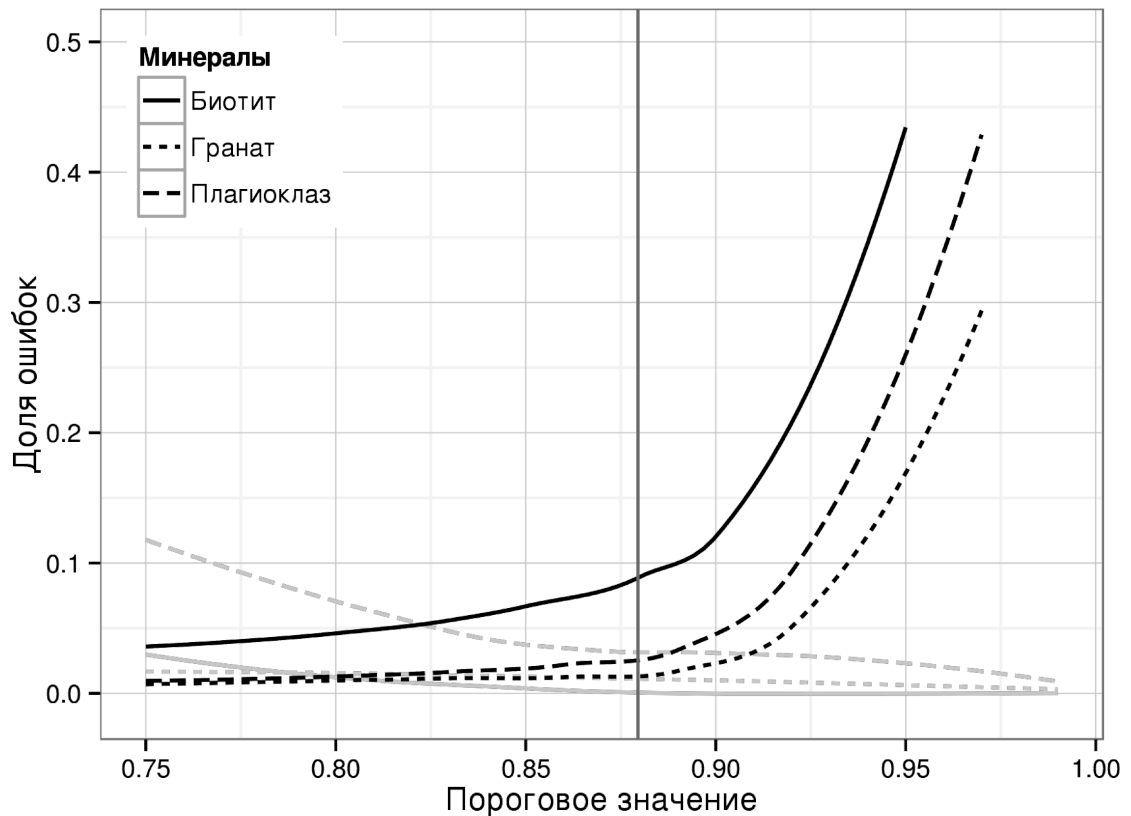


Рис. 4.1. Доля ошибок для рассматриваемых минералов при изменении порогового значения результирующей функции по [Иванов, 2016с]. Доля ложноположительных результатов показана серым, а ложноотрицательных — черным цветом. Вертикальной линией показано выбранное пороговое значение, равное 0.88

равным 0.5. Так как порядок следования элементов является наименее существенным из всех типов ограничений, значение τ не должно превосходить $1 - R$, поэтому оно было выбрано равным 0.1.

4.2. Программная реализация

В прототипе программы, на котором выполнялся подбор параметров, использовался построчный синтаксический разбор файлов с ограничениями. Этот подход был недостаточно быстрым в случае большого числа ограничений (тысячи минералов). Для решения этой проблемы был использован механизм трансляции с языка ограничений в исходный код на языке C++.

Полученный исходный код собирается компилятором в исполняемый модуль, получающий данные на стандартный ввод и выдающий результаты на стандартный вывод. Также выдается дополнительная информация,

которая может быть использована при отладке набора ограничений для минерала. Она выводится в формате с разделителями, предназначенном для дальнейшей автоматической обработки (листинг 4.2).

```
1 Name: 265-2_4 т.4
2 Mineral: Biotite
3 Variant: Dana, 71.02.02b: Biotite: K(Mg,Fe,Mn,Al,Ti)3 Al
  (AlSi) 0Si210 (F,OH)2: 0.166225
4 Formula: O:11.000000 Mg:1.689285 Al:1.402536 Si:2.729872
  K:0.916887 Ti:0.209649 Fe:0.860876 Norm: O: 11
5 Range: K:1 :0.166225
6 Range: Mg+Fe+Mn+Al+Ti:3~7 :0.000000
7 Range: Al:1~5 :0.000000
8 Range: Fe+Mg+Mn+Ti:0~3 :0.000000
9 Range: Si:2~3 :0.000000
10 Other: O, K, Mg, Mn, Fe, Ti, Al, Si :0.000777
```

Лист. 4.2. Вывод исполняемого модуля идентификации

Помимо состава образца программе также передается пороговое значение R , что позволяет пользователю регулировать точность определения. Вывод программы может включать несколько записей, аналогичных листингу 4.2 в случае, если подходящих минералов было найдено несколько. Помимо названия минерала вывод включает результаты пересчета (строка 4) и значение функций принадлежности для отдельных правил (строки 6-11). Использование такого подхода позволило добиться существенного повышения скорости идентификации и обеспечило возможность работы с многими тысячами минералов.

4.3. Автоматическое составление наборов ограничений

Ручное составление правил является трудоемкой задачей, поэтому актуальной проблемой становится автоматическая генерация правил на основе существующих минералогических справочников. Для этого на первом этапе были выбраны разделы, посвященные силикатам, из двух минералогических справочников Dana's New Mineralogy [Gaines et al., 1997] и Handbook of Mineralogy [Anthony et al., 1995]. В ходе автоматического анализа были получены названия минералов и их кристаллохимические формулы. Они были представлены в виде текстовых строк, например, $K(Mg, Fe, Mn, Al, Ti)3Al(AlSi)Si2O11+(H2O, F)$. Вторым этапом стало непосредственно составление наборов ограничений по этим формулам.

Однозначное преобразование такой формулы в набор ограничений, как правило, возможно, однако формула может быть недостаточно формализованной, например $(Mn^{+++}+1-x)(Mn^{++}x)(O2-2x)(OH2x)$ where $x = 0.06-0.07$. Другой проблемой автоматического преобразования является правильный расчет количества кислорода в формуле, содержащей несколько групп анионов. Поэтому при автоматическом составлении ограничений было введено два дополнительных условия. Во-первых, минералы с переменной в формуле отбрасывались, во-вторых, нормирование производилось не на кислород, как это принято для данного класса минералов, а на сумму кремния и алюминия в формуле, что является разновидностью пересчета на сумму катионов. Данный метод пересчета, как правило, не используется, так как эта сумма недостаточно стабильна, однако ее использование для идентификации допустимо.

К сожалению, рассмотренные минералогические справочники содержат только формулы конечных рядов твердых растворов, поэтому для многих породообразующих минералов наборы ограничений пока могут создаваться только вручную. Ручное создание таких наборов требует дополнительного тестирования. При этом приходится корректировать ограничения, полученные из исходной формулы минерала или вводить дополнительные. Трудности также вызывает составление наборов ограничений для минералов с очень развитым и сложным изоморфизмом, например, для группы амфиболов [Иванов, 2016с].

Тем не менее, в ходе автоматического составления набора ограничений, было получено в общей сложности 1791 набор.

4.4. Интеграция с интерактивным реестром

Для интеграции системы идентификации с интерактивным реестром были разработаны модуль сервера, выполняющий идентификацию, и модуль пользовательского интерфейса, отображающий результаты идентификации пользователю (рисунок 4.2). Также на модуль сервера была возложена задача трансляции правил и перекомпиляции исполняемых модулей при изменении набора правил на странице реестра. Перекомпиляция выполняется только при изменении содержания страницы. Этот процесс происходит непосредственно при выполнении идентификации, чтобы

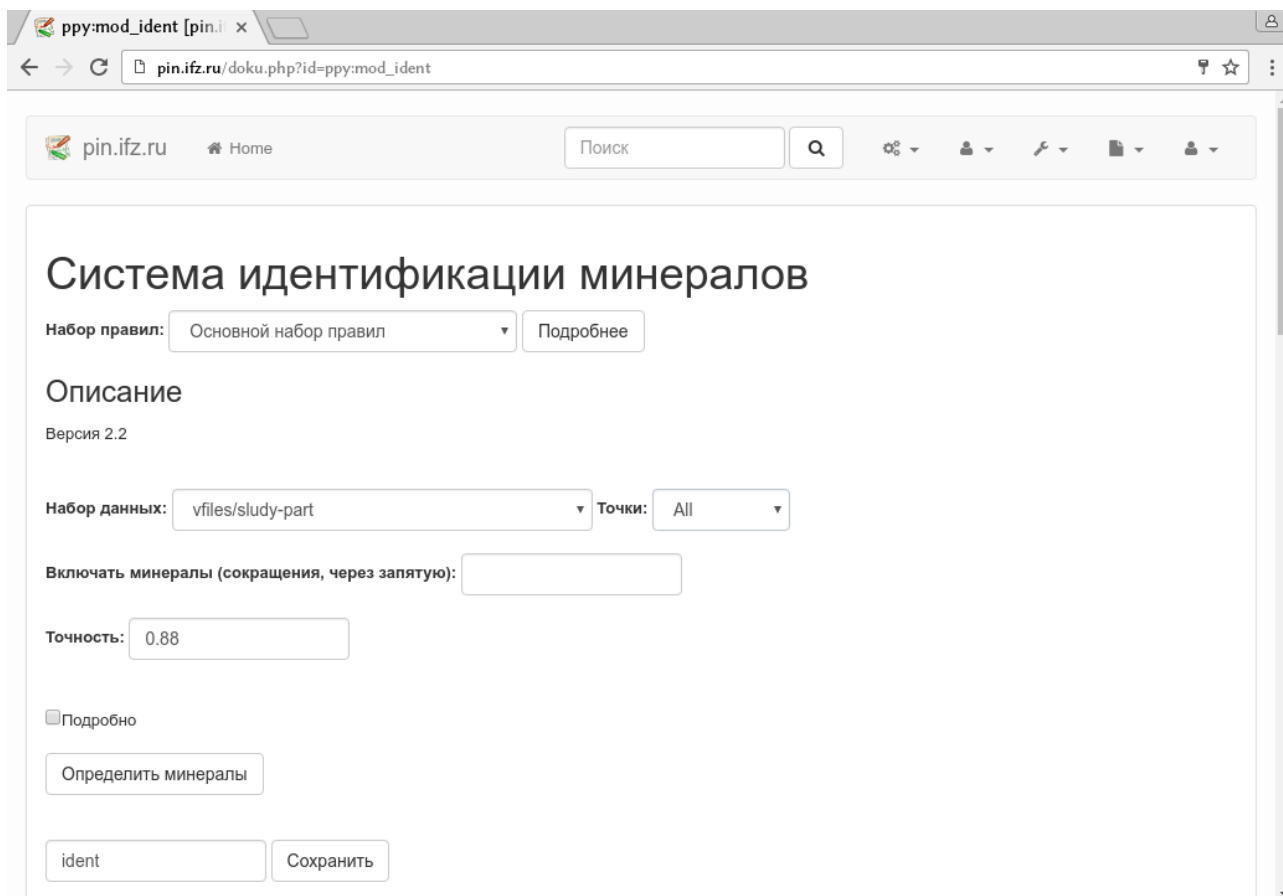


Рис. 4.2. Интерфейс системы идентификации минералов на базе интерактивного реестра

избежать повышенной нагрузки на сервер при неоднократных правках страницы с набором правил.

Способ описания ограничений на составы минералов в системе идентификации достаточно прост, что позволяет пользователю как создавать свои наборы для решения специализированных задач, так и изучать уже существующие. Наборы ограничений, полученные в результате автоматической обработки, доступны только в виде предварительно собранных исполняемых модулей в связи с тем, что их перекомпиляция создает слишком большую дополнительную нагрузку на сервер. Кроме того, отдельная страница, содержащая нескольких тысяч наборов ограничений, неудобна для работы с ней в рамках вики.

На созданное приложение для идентификации минералов получено авторское свидетельство [Свидетельство ...№ 2016616340, Иванов, 2016].

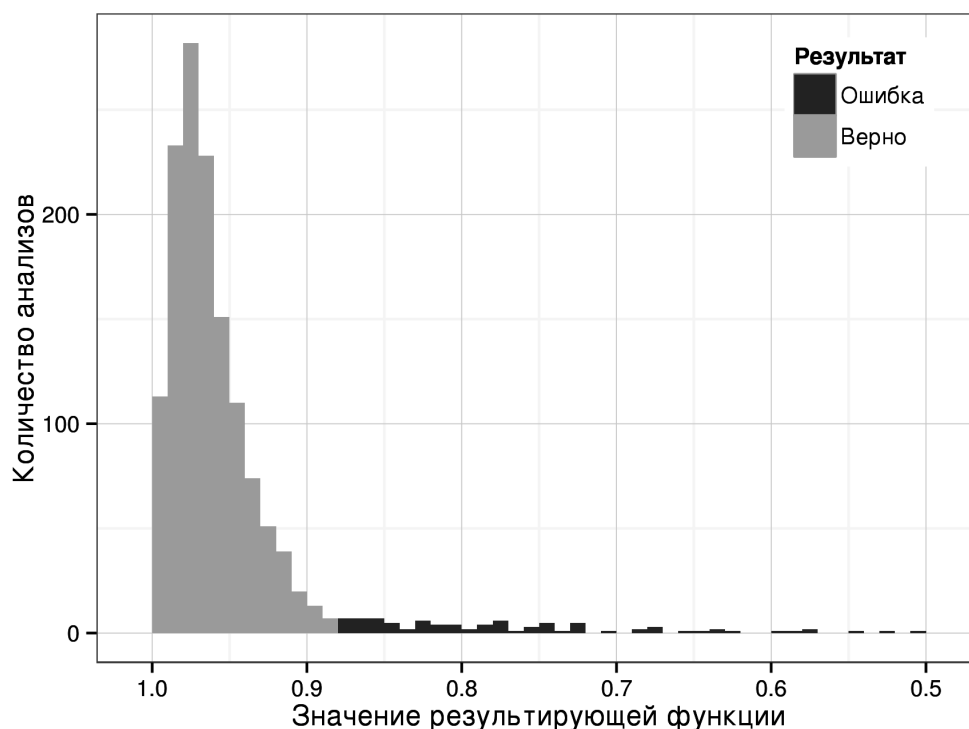


Рис. 4.3. Распределение значений результирующей функции и их интерпретация в соответствии с выбранным пороговым значением

4.5. Использование системы

Для апробации системы был использован весь предоставленный А. Л. Кулаковским набор результатов микрозондовых измерений. Для всех минералов и твердых растворов, представленных в наборе данных (гранат, биотит, хлорит, корунд, клинопироксен, кордиерит, гиперстен, полевой шпат, мусковит, ставролит, шпинель и еще двадцать два распространенных порообразующих минерала и группы), были вручную заданы наборы ограничений. Значения результирующей функции и их интерпретация представлены на рисунке 4.3. Из 1428 анализов согласуются с экспертными оценками 1301 (91 %) [Иванов, 2016с].

Также система была применена для отбора исходных данных в работе, посвященной фенгитовой барометрии [Камзолкин, Иванов, Конилов, 2015]. Для этого был создан набор ограничений для выделения фенгита из непрерывного мусковит-селадонитового ряда (листинг 4.3), что позволило выполнить отбор проб для дальнейшего использования в геотермобарометрии (см. раздел 3.4).

```
1 mineral: Phe
2 name: Phengite
3 id: name, Phengite
4 formula: K(Al,Mg,Fe)2(Si,AlO)410(OH)2
5 norm: 0,11
6 elements: O,K,Al,Mg,Si,Fe
7 rule: K= 0.8~1
8 rule: Al,Mg,Fe=2~8
9 rule: Mg,Fe=0~2
10 rule: Al=0~6
11 rule: Si,Al=4~6
```

Лист. 4.3. Набор правил для определения фенгита

4.6. Выводы

В данной главе рассмотрено построение системы идентификации минералов по химическому составу. Она использует механизм на основе нечеткой логики. Для настройки параметров системы было выполнено тестирование на большом объеме геохимических данных.

Разработанная система была интегрирована с интерактивным реестром, причем архитектура реестра позволяет использовать многочисленные наборы правил, выделенные в отдельные страницы. Такие наборы могут быть как специализированными, так и общего назначения.

Предпринятая попытка автоматической генерации правил по кристаллохимическим формулам, представленным в литературе, позволила получить большое количество наборов ограничений. Однако недостаточная степень их формализации не позволяет использовать эти наборы в полной мере. Прежде всего проблема связана тем, что в минералогической справочной литературе информация представлена преимущественно для конечных членов непрерывных рядов твердых растворов. Другая проблема связана с расчетом числа атомов кислорода, на которое производится нормирование.

Применение системы к используемой коллекции геохимических данных позволило добиться высокого процента успешных определений, однако для этого потребовалось ручное составление и тестирование правил.

Основными направлениями использования разработанной системы являются контроль и отбор результатов химического анализа горных пород для дальнейшего использования в расчетах, например, в геобаротермометрии.

Как правило название минерала известно. Поэтому определение неизвестного заранее минерала является вторичной задачей по сравнению с проверкой соответствия между известным названием и полученным составом.

В форме наборов ограничений для данной системы могут выступать, например, требования к составу образцов, налагаемые геосенсорами. Использование системы в качестве «входного фильтра» позволит обеспечить бóльшую достоверность результатов, полученных методами геобаротермометрии.

Включение в реестр системы идентификации минералов демонстрирует возможность его использования для построения систем на основе декларативных языков.

Заключение

В ходе данной работы было создано веб-приложение — реестр методов обработки петролого-геохимических данных. Его основными отличиями от существующих на сегодняшний день являются: формат веб-приложения, сравнительно новый для научных приложений; использование вики-системы в качестве платформы для построения реестра; открытость исходного кода методов и возможность для пользователей не только использовать имеющиеся методы, но и непосредственно участвовать в развитии реестра, добавляя методы и улучшая документацию к существующим.

В результате анализа существующих решений в области обработки петролого-геохимических данных (глава 1) был составлен обзор существующих инструментов. В него вошли приложения для организации и управления данными, построения геохимических диаграмм, идентификации минералов, геобаротермометрии и др. Были выявлены недостатки, присущие существующим инструментам.

Сформулированы основные принципы, которые легли в основу построения будущего интерактивного реестра (глава 2). В качестве платформы для построения реестра была выбрана система на основе технологии вики. Основные операции в пользовательском интерфейсе реестра представлены в приложении 1. Ключевой технической задачей в рамках данной работы стала реализация механизма, который использует для обработки данных алгоритм, находящийся на странице вики. Для этого был создан промежуточный сервер, взявший на себя управление вычислениями и импорт кода из вики-системы.

На базе интерактивного реестра создано несколько модулей для решения прикладных задач. Модули минералогических пересчетов и интерпретации составов минералов предназначены для упрощения рутинных задач обработки петролого-геохимическими данными. Модуль геобаротермометрии позволяет выполнять оценки давления и температуры с использованием геосенсоров, имеющихся в реестре. Созданы модули для решения ряда задач, связанных с визуализацией геохимических данных. Кроме того, разработана система идентификации, позволяющая выполнять

определение наименований минералов по их химическому составу с помощью имеющихся в реестре наборов ограничений.

Использование реестра позволило получить новые научные данные в исследованиях истории блыбского комплекса Передового хребта Большого Кавказа.

Для дальнейшего развития реестра необходимо привлечение специалистов — петрологов и геохимиков — к расширению реестра и документированию имеющихся в нем методов, а также к получению отзывов по использованию для дальнейшего совершенствования интерфейса и расширения возможностей. По мере развития реестра предполагается все большая формализация информации о методах с одной стороны и рост объема неформализуемой текстовой информации с другой. При этом становится возможно сформулировать формальные критерии качества для самих инструментов и их документации. По мере роста информационной наполненности реестра, открываются возможности по его использованию не только в науке, он и в образовании в качестве справочного материала и учебного инструмента.

После обеспечения достаточного покрытия методов тестами возможен их экспорт в виде библиотеки в репозиторий CRAN для повышения доступности методов для конечных пользователей.

Использование сформулированной в рамках данной работы концепции возможно не только для решения задач, связанных с обработкой геохимических данных. Можно рассмотреть использование такого подхода для любых задач, характеризующихся большим количеством методов обработки достаточно стандартных наборов данных. Причем, возможными приложениями системы видятся не только научные, но и прикладные задачи. Например, создание и заполнение форм документов, генерация кода для начального конфигурирования микроконтроллеров и других устройств и др.

Если говорить о далеко идущих перспективах, то создание реестра может стать первым этапом построения экспертной системы для работы с геохимическими данными в помощь геологу. Однако её создание потребует интеллектуального анализа огромного объема текстовых данных для построения системы типа вопрос-ответ. Примером такой системы является IBM DeepAQ [Ferrucci et al., 2010].

Литература

1. Авченко О.В. и др. Генезис гранатосодержащих пород Березитового месторождения (Верхнее Приамурье, Россия) // Геология рудных месторождений. — 2014. — Т. 56, № 1. — С. 19–40.
2. Аранович Л.Я. Минеральные равновесия многокомпонентных твердых растворов. — М.: Наука, 1991. — С. 253.
3. Дубровский М.И. CIPWD — новый алгоритм пересчета результатов химического анализа на минералы // Вестн. Воронеж. ун-та, Геология. — 2000. — Т. 5, № 10. — С. 98–114.
4. Заде Л.А. Понятие лингвистической переменной и его применение к принятию приближенных решений / пер. с англ. Н.И. Ринго. — М.: Мир, 1976. — С. 165.
5. Иванов С.Д. Современная платформа для обработки результатов геохимических анализов // Физико-химические и петрофизические исследования в науках о Земле. Материалы XV международной конференции. — М.: ГЕОХИ, 2014а. — С. 92–94.
6. Иванов С.Д. Экспертная система определения минералов по результатам микрозондовых анализов // Конференция молодых ученых ИФЗ РАН. — Москва: ИФЗ РАН, 2014б. — С. 44.
7. Иванов С.Д. Интерактивная база методов петрохимической обработки данных на основе технологии вики на примере системы идентификации минералов // Конференция молодых ученых ИФЗ РАН. — Москва: ИФЗ РАН, 2015а.
8. Иванов С.Д. Некоторые требования к интерактивной базе методов обработки геохимических данных // Науки о Земле. Современное состояние. Материалы третьей всероссийской молодежной научно-практической школы-конференции. — Шира: НГУ, 2015б. — С. 119–120.
9. Иванов С.Д. Интерактивный реестр геосенсоров на основе веб-приложения // Компьютерные исследования и моделирование. — 2016а. — Т. 8, № 4. — С. 621–632.
10. Иванов С.Д. Интерактивный реестр методов обработки петролого-геохимических данных // Научная конференция молодых ученых и аспирантов ИФЗ РАН. Тезисы докладов и программа Конференции. — Москва: ИФЗ РАН, 2016б. — С. 33.

11. *Иванов С.Д.* Метод идентификации минералов по химическому составу на основе нечеткой логики // Геофизические исследования. — 2016с. — Т. 17, № 4. — С. 47–56. — DOI: 10.21455/gr2016.4-4.
12. *Камзолкин В.А.* — Тектоника домезозойского основания западной части Передового хребта Северного Кавказа [Текст]. — Диссертация ... кандидата геол.-мин. наук. М.: ИФЗ РАН, 2013. — 186 с.
13. *Камзолкин В.А., Иванов С.Д., Конилов А.Н.* Некоторые результаты анализа зависимости состава белых слюд (фенгита) от РТ-условий их формирования // Физико-химические и петрофизические исследования в науках о Земле. Материалы XIV международной конференции. — М.: ГЕОХИ, 2013. — С. 123–126.
14. *Камзолкин В.А., Иванов С.Д., Конилов А.Н.* Эмпирический фенгитовый геобарометр: Обоснование, калибровка и применение // Записки Российского Минералогического Общества. — 2015. — Т. 144, № 5. — С. 1–14.
15. *Кориневский Е.В.* PetroExplorer – система для создания геохимических информационно-аналитических массивов в процессе тематических исследований // Геоинформатика. — 2015. — Т. 4. — С. 48–53.
16. *Перчук А.Л.* — Петрология и минеральная хронометрия коровых эклогитов [Текст]. — Диссертация ... доктора геол.-мин. наук. М.: МГУ, 2004. — 333 с.
17. *Перчук Л.Л., Рябчиков И.Д.* Фазовое соответствие в минеральных системах. — М.: Недра, 1976. — С. 287.
18. *Петров Т.Г.* Информационный язык РНА для описания, систематизации и изучения изменений составов многокомпонентных объектов // Научно-техническая информация. — 2001. — Т. 3. — С. 8–18.
19. Петрология метаморфических комплексов Большого Кавказа / под ред. С. Кориковский. — М.: Наука, 1991. — С. 231.
20. *Свидетельство о государственной регистрации программы для ЭВМ № 2016616340 Российская Федерация.* Программа для идентификации минералов по химическому составу на основе нечеткой логики / С.Д. Иванов ; заявитель и правообладатель ИФЗ РАН. — № 2016613557 ; заявл. 13.04.2016 ; опубл. 09.05.2016.
21. *Сизых А.И., Буланов В.А.* Термометрия, барометрия, петрохимия магматических и метаморфических пород. — Иркутск: Изд-во Иркут. ун-та, 1991. — С. 232.

22. *Симакин А.Г., Закревская О.Ю., Салова Т.П.* Оценка условий кристаллизации кортландитов Камчатки по составам амфиболов // Вестник Отделения наук о Земле РАН. — 2012. — Т. 4. — С. 1–4.
23. *Скляр Е.В.* Интерпретация геохимических данных: Учеб. пособие. — М.: Интернет Инжиниринг, 2001. — С. 288.
24. *Скублов С.Г.* Геохимия редкоземельных элементов в породообразующих метаморфических минералах. — СПб.: Наука, 2005. — С. 147.
25. *Чудненко К.В., Авченко О.В., Вах А.С.* Программа MC — петрологический инструмент для вычисления реальных количеств минералов в горной породе. — 2013. — URL: http://fegi.ru/elibrary/elibrary/doc_details/361 (дата обр. 14.03.2016).
26. *Anderson E. et al.* LAPACK Users' Guide (Third Ed.) — Philadelphia, USA: Society for Industrial, Applied Mathematics, 1999. — P. 404. — ISBN 0-89871-447-8.
27. *Andreadisa G., Fourtounisb G., Bouzakisa K.-D.* Collaborative design in the era of cloud computing // Advances in Engineering Software. — 2015. — Vol. 81. — Pp. 66–72.
28. *Anthony J.W. et al.* Handbook of Mineralogy, Volume 2: Silica, Silicates. — Chantilly, France: Mineral Data Publishing, 1995. — P. 904.
29. AppArmor Documentation. — 2016. — URL: <http://wiki.apparmor.net/index.php/Documentation> (visited on 07/14/2016).
30. *Barrett D.J.* MediaWiki. — Sebastopol, USA: O'Reilly Media, 2009. — P. 361.
31. *Beck K.* Test-driven development: by example. — London, UK: Addison-Wesley Professional, 2003. — P. 220.
32. *Berman R.G.* winTWQ (version 2.3): a software package for performing internally-consistent thermobarometric calculations // Geological Survey of Canada. — 2007.
33. *Bernhardt H.-J.* MINCALC-V5 a non EXCEL based computer program for general electron-microprobe mineral analyses data processing // Abstract of 20th IMA-Meeting. — Budapest, Hungary: IMA, 2010. — P. 869.
34. *Brandelik A., Massonne H.-J.* PTGIBBS — an EXCEL™ Visual Basic program for computing and visualizing thermodynamic functions and equilibria of rock-forming minerals // Computers & Geosciences. — 2004. — Vol. 30, No. 9–10. — Pp. 909–923.

35. *Brey G.P., Bulatov V.K., Girnits A.V.* Geobarometry for Peridotites: Experiments in Simple and Natural Systems from 6 to 10 GPa // *Journal of Petrology*. — 2008. — Vol. 49, No. 1. — Pp. 3–24.
36. *Cabanis B.* Le diagramme La/10-Y/15-Nb/8 : unoutil pour la discrimination des series volcaniques et la mise en evidence des processus de melange et/ou de contamination crustale // *C. R. Acad. Sci. Ser. II*. — 1989. — T. 309. — C. 2023–2029. — URL: <http://ci.nii.ac.jp/naid/80004995562/en/>.
37. *Caddick M.J., Thompson A.B.* Quantifying the tectono-metamorphic evolution of pelitic rocks from a wide range of tectonic settings: mineral compositions in equilibrium // *Contrib. Mineral. Petrol.* — 2008. — Vol. 156. — Pp. 177–195.
38. *Capitani C. de, Petrakakis K.* The computation of equilibrium assemblage diagrams with Theriak/Domino software // *American Mineralogist*. — 2010. — Vol. 95. — Pp. 1006–1016.
39. *Caritat P. de, Bloch J., Hutcheon I.* LPNORM: A linear programming normative analysis code // *Computers & Geosciences*. — 1994. — Vol. 3, No. 20. — Pp. 313–347.
40. *Carr M.J.* Igpert for Windows. — Somerset: Terra Softa Inc, 2002. — P. 38.
41. *Chambers J.A., Kohn M.J.* Titanium in muscovite, biotite, and hornblende: Modeling, thermometry, and rutile activities of metapelites and amphibolites // *Amer. Miner.* — 2012. — Vol. 97. — Pp. 543–555.
42. *Coggon R., Holland T.J.B.* Mixing properties of phengitic micas and revised garnet-phengite thermobarometers // *J. Metamorphic Geol.* — 2002. — Vol. 20, No. 7. — Pp. 683–696.
43. *Cohen D., Ward C.R.* SEDNORM — a program to calculate a normative mineralogy for sedimentary rocks based on chemical analyses // *Computers & Geosciences*. — 1991. — Vol. 17, No. 9. — Pp. 1235–1253.
44. *Collins-Sussman B., Fitzpatrick B., Pilato M.* Version Control with Subversion. — Sebastopol, USA: O'Reilly Media, Inc., 2004. — P. 299.
45. *Connolly J.A.D.* The geodynamic equation of state: what and how // *Geochemistry, Geophysics, Geosystems*. — 2009. — Vol. 10. — Q10014.
46. *Dachs E.* PET: petrological elementary tools for Mathematica // *Computers & Geosciences*. — 1998. — Vol. 24, No. 3. — Pp. 219–235.
47. *Eckert J.O., Newton R.C., Kleppa O.J.* The H of reaction and recalibration of garnet-pyroxene-plagioclase-quartz geobarometers in the

- CMAS system by solution calorimetry // *American Mineralogist*. — 1991. — T. 76, № 1–2. — C. 148–160.
48. *Eisenberg J.D., Bellamy-Royds A.* SVG essentials. Second Edition. — Sebastopol, USA: O'Reilly Media, Inc., 2014. — P. 345.
 49. *Ellis D.J., Green D.H.* An experimental study of the effect of Ca upon garnet-clinopyroxene Fe-Mg exchange equilibria // *Contributions to Mineralogy and Petrology*. — 1979. — No. 71. — Pp. 13–22.
 50. *Ferrucci D. et al.* Building Watson: An overview of the DeepQA project // *AI magazine*. — 2010. — Vol. 31, No. 3. — Pp. 59–79.
 51. *Fielding R.T.* — Architectural Styles and the Design of Network-based Software Architectures, dissertation, submitted in partial satisfaction of the requirements for the degree of doctor of philosophy, in Information and Computer Science. — Master's thesis. Irvine, USA: University of California, 2000. — P. 180.
 52. *Folorunso I.O. et al.* A Rule-Based Expert System for Mineral Identification // *Journal of Emerging Trends in Computing and Information Sciences*. — 2012. — Vol. 3, No. 2. — Pp. 205–210.
 53. *Fonarev V.I., Graphchikov A.A., Konilov A.N.* A consistent system of geothermometers for metamorphic complexes // *International Geology Review*. — 1991. — Vol. 33, No. 8. — Pp. 743–783.
 54. *Gaines R.V. et al.* Dana's New Mineralogy: The System of Mineralogy of James Dwight Dana and Edward Salisbury Dana. 8th Edition. — Hoboken, USA: Wiley-Interscience, 1997. — P. 1872.
 55. Geochron. — 2011. — URL: <http://www.geosamples.org/> (visited on 09/14/2016).
 56. GEOROC database (Geochemistry of Rocks of the Oceans and Continents). — 2004. — URL: <http://georoc.mpch-mainz.gwdg.de/georoc/> (visited on 09/14/2016).
 57. *Ghiorso M.S., Kress V.C.* Computational Thermodynamics Server. — 2011. — URL: <http://ctserver.ofm-research.org/geothermometer.html> (visited on 06/02/2016).
 58. *Gyerik J.* Bazaar Version Control. — Birmingham, UK: Packt Publishing, 2013.
 59. *Haneberg W.* Computational geosciences with Mathematica. — Seattle, USA: Springer Science & Business Media, 2012. — P. 363.

60. *Heesch D. van Doxygen* — Generate documentation from source code. — 1997. — URL: <http://www.stack.nl/~dimitri/doxygen/index.html> (visited on 07/14/2016).
61. *Holland T., Blundy J.* Non-ideal interactions in calcic amphiboles and their bearing on amphibole-plagioclase thermometry // *Contributions Mineral Petrology*. — 1994. — Vol. 116. — Pp. 433–447.
62. *Hora J.M. et al.* An Excel-based tool for evaluating and visualizing geothermobarometry data // *Computer & Geosciences*. — 2013. — Vol. 56. — Pp. 178–185.
63. *Hunt B.R., Lipsman R.L., Rosenberg J.M.* A guide to MATLAB: for beginners and experienced users. — Cambridge, UK: Cambridge University Press, 2014. — P. 310.
64. *Ince D.C., Hatton L., Graham-Cumming J.* The case for open computer programs // *Nature*. — 2012. — Vol. 482, No. 7386. — Pp. 485–488.
65. *Ivanov S.D. et al.* Chemical and mineralogical data and processing methods management system prototype with application to study of the North Caucasus Blybsky Metamorphic Complexes metamorphism PT-condition // *EGU General Assembly 2014, Geophysical Research Abstracts*. Vol. 16. — Viene: EGU, 2014. — P. 845.
66. *Janousek V.* Interpretation of Whole-rock Geochemical Data in Igneous Geochemistry: Introducing Geochemical Data Toolkit (GCDkit) // *Journal of Petrology*. — 2006. — Vol. 47, No. 6. — Pp. 1255–1259.
67. javadoc - The Java API Documentation Generator. — 2004. — URL: <http://docs.oracle.com/javase/1.5.0/docs/tooldocs/solaris/javadoc.html> (visited on 07/14/2016).
68. *Kabacoff R.I.* R in Action, Second Edition. Data analysis and graphics with R. — New York, USA: Manning Publications, 2015. — P. 608.
69. *Kelsey C.H.* Calculation of the CIPW norm // *Mineralogical Magazine*. — 1965. — No. 34. — Pp. 276–282.
70. *Khattree R., Naik D.N.* Andrews plots for multivariate data: some new suggestions and applications // *Journal of Statistical Planning and Inference*. — 2002. — Vol. 100, No. 2. — Pp. 411–425.
71. *Kroah-Hartman G.* Linux Kernel in a Nutshell. — Sebastopol, USA: O'Reilly Media, 2006. — P. 182.
72. *Laboratories B.T., Kernighan B., McIlroy M.* UNIX programmer's manual. — Murray Hill, New Jersey, USA: Bell Telephone Laboratories, Incorporated, 1979. — P. 404.

73. *Leuf B.A., Cunningham W.* The Wiki Way: Quick Collaboration on the Web. — Boston, USA: Addison-Wesley, 2001. — P. 362.
74. Library of Experimental Phase Relations. — 2008. — URL: <http://lepr.ofm-research.org/> (visited on 09/14/2016).
75. *Loeliger J., McCullough M.* Version Control with Git. — Sebastopol, USA: O'Reilly Media, Inc., 2012. — P. 434.
76. *MacGregor I.D.* Empirical geothermometers and geothermobarometers for spinel peridotite phase assemblages // International Geology Review. — 2015a. — Vol. 57, No. 15. — Pp. 1940–1974.
77. *MacGregor I.D.* Empirical geothermometers and geothermobarometers for spinel peridotite phase assemblages // International Geology Review. — 2015b. — Vol. 57, No. 15. — Pp. 1940–1974.
78. *Massonne H.J., Schreyer W.* Phengite geobarometry based on the limiting assemblage with K-feldspar, phlogopite, and quartz // Contrib. Mineral. Petrol. — 1987. — Vol. 96. — Pp. 212–224.
79. *Massonne H.J., Szpurka Z.* Thermodynamic properties of white micas on the basis of high-pressure experiments in the systems K₂O-MgO-Al₂O₃-SiO₂-H₂O and K₂O-FeO-Al₂O₃-SiO₂-H₂O // Lithos. — 1997. — Vol. 41. — Pp. 229–250.
80. *McDonough W.F., Sun S.-s.* The composition of the Earth // Chem. Geol. — 1995. — T. 120. — C. 223–253.
81. *McGuire P.* Getting Started with Pyparsing. — Sebastopol, USA: O'Reilly Media, 2008. — P. 65.
82. *Miyashiro A., Shido F.* Tschermak Substitution in Low- and Middle-grade Pelitic Schists // Journal of Petrology. — 1985. — T. 26. — C. 449–487.
83. *Moutte J.* arxim, a library for thermodynamic modeling of fluid-rock systems. — Saint-Étienne: Ecole des Mines de Saint Etienne, 2009. — P. 55.
84. *Nakamura D.* A new formulation of garnet-clinopyroxene geothermometer based on accumulation and statistical analysis of a large experimental data set // Journal of Metamorphic Geology. — 2009. — Vol. 27, No. 7. — Pp. 495–508.
85. *Nakamura D.* Excel sheets for thermodynamic calculations. — 2015. — URL: <http://www.sky.sannet.ne.jp/minadai/ToolE.htm> (visited on 03/14/2016).
86. NAVDAT, The North American Volcanic and Intrusive Rock Database. — 2014. — URL: <http://www.navdat.org/> (visited on 09/14/2016).

87. *Omoumi H.* — GemIdent: a data base for gems and some applications of the electron microprobe in gem characterisation. — Master's thesis. Alberta: University of Alberta, 1990. — P. 270.
88. *O'reilly T.* What is Web 2.0: Design patterns and business models for the next generation of software // Communications & strategies. — 2007. — No. 1. — P. 17.
89. *Owen J.V., Greenough J.D.* An empirical sapphirine-spinel MgFe exchange thermometer and its application to high grade xenoliths in the Popes Harbour dyke, Nova Scotia, Canada // Lithos. — 1991. — Vol. 26, No. 3. — Pp. 317–332.
90. *Paktunc A.D.* MODAN — a computer program for estimating mineral quantities based on bulk composition: windows version // Computers & Geosciences. — 2001. — Vol. 27, No. 7. — Pp. 883–886.
91. *Perchuk A.L., Philippot P.* Rapid cooling and exhumation of eclogitic rocks from the Great Caucasus, Russia // J. Metamorph. Geol. — 1997. — T. 15. — C. 299–310.
92. *Perkins D., Chipera S.J.* Garnet-orthopyroxene-plagioclase-quartz barometry: refinement and application to the English River subprovince and the Minnesota River valley // Contributions to Mineralogy and Petrology. — 1985. — T. 89, № 1. — C. 69–80.
93. *Perroud P.* ATHENA — Mineral database. — 2013. — URL: <http://athena.unige.ch/athena/mineral/mineral.html> (visited on 04/14/2016).
94. PetDB, the Petrological Database. — 2012. — URL: <http://www.earthchem.org/petdb> (visited on 09/14/2016).
95. *Petrelli M. et al.* Petrograph: a New Software to Visualize, Model, and Present Geochemical Data in Igneous Petrology // Geochemistry, Geophysics, Geosystems. — 2005. — Vol. 6. — Q07011.
96. *Plante A., Peck D., Bargaen D.V.* The Mineral Identification Key II. — 2003. — URL: http://www.minsocam.org/msa/collectors_corner/id/mineral_id_keyi1.htm (visited on 07/14/2016).
97. *Posch M., Kurz D.* A2M — A program to compute all possible mineral modes from geochemical analyses // Computers & Geosciences. — 2007. — Vol. 33, No. 4. — Pp. 563–572.
98. *Powell R., Holland T.J.B., Worley B.* Calculating phase diagrams involving solid solutions via non-linear equations, with examples using

- THERMOCALC // Journal of Metamorphic Geology. — 1998. — Vol. 16. — Pp. 577–588.
99. *Preston J.* Gabbrosoft — Mineralogical spreadsheet. — 2011. — URL: <http://www.gabbrosoft.org/spreadsheets.html> (visited on 03/14/2016).
 100. *Pride C., Muecke G.K.* Rare earth element distributions among coexisting granulite facies minerals, Scourian complex NW Scotland // Contrib. Mineral. Petrol. — 1981. — T. 76. — C. 463–471.
 101. *Prlić A., Procter J.B.* Ten Simple Rules for the Open Development of Scientific Software // PLOS Computational Biology. — 2012. — Vol. 12, No. 8. — e1002802.
 102. *Putirka K.D.* Thermometers and Barometers for Volcanic Systems // Reviews in Mineralogy and Geochemistry. — 2008. — Vol. 69, No. 1. — Pp. 61–120.
 103. *Rameshwar D.R.* BGT — The macros driven spreadsheet program for biotite-garnet thermometry // Computer & Geosciences. — 1995. — Vol. 21, No. 4. — Pp. 593–604.
 104. *Ravna K.* The garnet-clinopyroxene Fe²⁺-Mg geothermometer: an updated calibration // Journal of Metamorphic Geology. — 2000. — Mar. — Vol. 18, No. 2. — Pp. 211–219. — DOI: 10.1046/j.1525-1314.2000.00247.x.
 105. *Reche J., Martinez F.J.* GPT: An Excel spreadsheet for thermobarometric calculations in metapelitic rocks // Computer & Geosciences. — 1996. — Vol. 22, No. 7. — Pp. 774–784.
 106. *Reed S.J.B.* Electron Microprobe Analysis and Scanning Electron Microscopy in Geology. Second Edition. — Cambridge, UK: Cambridge University Press, 2005. — P. 216.
 107. *Reenskaug T.* The Model-View-Controller (MVC). Its Past and Present // JavaZONE. — Oslo, Norway, 2003. — P. 16.
 108. *Rhodes B., Goerzen J.* Foundations of Python network programming. — New York, USA: Apress, 2010. — P. 345.
 109. *Richard L.R.* MinPet: Mineralogical and Petrological Data Processing System, Version 2.02. — Québec: MinPet Geological Software, 1995.
 110. *Riefstahl F.* FastGAPP: A user-friendly MATLAB-based program helps geoscientists classifying, interpreting and displaying igneous rocks in terms of petrography and geochemistry. — Bremen: University of Bremen, 2012. — P. 32.

111. *Rosen O.M., Abbyasov A.A., Tipper J.C.* MINLITH — an experience-based algorithm for estimating the likely mineralogical compositions of sedimentary rocks from bulk chemical analyses // *Computers & Geosciences*. — 2004. — Vol. 30, No. 6. — Pp. 647–661.
112. SedDB. — 2013. — URL: <http://www.earthchem.org/seddb> (visited on 09/14/2016).
113. *Shafranovich Y.* RFC4180: Common Format and MIME Type for Comma-Separated Values (CSV) Files // *The Internet Society*. — 2005. — P. 6.
114. Shiny — A web application framework for R. — 2016. — URL: <http://shiny.rstudio.com/> (visited on 07/14/2016).
115. *Smith D.G.W., Leibovitz D.P.* MinIdent: A data base for minerals and a computer program for their identification // *Canadian Mineralogist*. — 1986. — Vol. 24, No. 4. — Pp. 695–708.
116. *Spear F.S.* Metamorphic Phase Equilibria and Pressure-Temperature-Time Paths. — Washington, USA: Mineral Society of America, 1995. — P. 799.
117. *Starek M. et al.* An Expert System for Mineral Identification // *J. Comput. Sci. Coll.* — 2002. — Vol. 17, No. 5. — Pp. 193–197.
118. *Strekopytov S.* REE distribution in recent pacific sediments and the relationship with major elements geochemistry and source provenances // *J. Czech Geol. Soc.* — 1997. — T. 42. — C. 48.
119. *Sturm R.* PX-NOM an interactive spreadsheet program for the computation of pyroxene analyses derived from the electron microprobe // *Computers & Geosciences*. — 2002. — Vol. 28, No. 4. — Pp. 473–483.
120. *Su Y., Langmuir C.H., Asimow P.D.* PetroPlot: A plotting and data management tool set for Microsoft Excel // *Geochemistry, Geophysics, Geosystems*. — 2003. — Vol. 4, No. 3. — GC000323.
121. *Sun S.S., McDonough W.F.* Chemical and isotopic systematics of oceanic basalts: implications for mantle composition and processes // *Geological Society, London, Special Publications*. — 1989. — T. 42, No 1. — C. 313–345.
122. *Swartz A.* web.py — Get Started. — 2016. — URL: <http://webpy.org/> (visited on 07/14/2016).
123. The TIOBE Programming Community index. — 2016. — URL: <http://www.tiobe.com/tiobe-index/> (visited on 07/14/2016).
124. *Tindle A.* Mineral recalculation software. — 2015. — URL: <http://www.open.ac.uk/earthresearch/tindle/AGTWebPages/AGTSoft.html> (visited on 05/16/2016).

125. *Tischendorf G. u dp.* True and brittle micas: composition and solid-solution series // *Mineral. Mag.* — 2007. — T. 71, № 10. — C. 285–320.
126. *Urbanek S.* Rserve — a fast way to provide R functionality to applications // *Proceedings of the 3rd International Workshop on Distributed Statistical Computing.* — Vienna, Austria, 2003. — P. 11.
127. *Victorino T.J.* Dokuwiki. — Saarbrücken, Germany: Log Press, 2012. — P. 100.
128. *Waters D.* Practical Aspects of Mineral Thermobarometry. Tools and accessories. — 2004. — URL: https://www.earth.ox.ac.uk/~davewa/pt/th_tools.html (visited on 06/02/2016).
129. *Wickham H.* ggplot2: Elegant Graphics for Data Analysis. — New York, USA: Springer, 2009. — P. 212.
130. *Wilkinson L.* The Grammar of Graphics. Second Edition. — New York, USA: Springer Science & Business Media, 2006. — P. 690.
131. *Yavuz F.* WinPyrox: A Windows program for pyroxene calculation classification and thermobarometry // *American Mineralogist.* — 2013. — Vol. 98, No. 7. — Pp. 1338–1359.

Список иллюстраций

1.1	Интерфейс приложений <i>PetroGraph</i> (а) и <i>PetroExplorer</i> (б) . . .	15
1.2	Интерфейс GCDKit (из статьи [Janousek, 2006])	17
1.3	Интерфейс программы TPF	22
2.1	Пример документация, сгенерированной системой Doxygen для библиотеки D-BUS	30
2.2	Два различных подхода к организации систем управления содержанием (CMS)	32
2.3	Структура интерактивного реестра	36
2.4	Программная архитектура интерактивного реестра	41
2.5	Файловый менеджер elFinder интегрированный с DokuWiki . . .	43
2.6	Интерфейс редактирования страницы системы DokuWiki	50
2.7	Интерфейс модуля минералогических пересчетов	52
2.8	Пример тригонограммы железо-цинк-серебро	55
2.9	Интерфейс построения диаграмм	55
2.10	Пример диаграммы Эндрюса [Khattree, Naik, 2002] для составов различных минералов	56
2.11	Пример дискриминационной диаграммы	58
2.12	Пример спайдерграммы	58
3.1	Пример страницы геосенсора	68
3.2	Интерфейс модуля геотермобарометрии	69
3.3	Три группы образцов фенгита в зависимости от содержания кремния и температуры	74
3.4	Сравнение абсолютной ошибки определения давления ($P - P_{calc}$) с использованием фенгитовых мономинеральных барометров	75
4.1	Доля ошибок для рассматриваемых минералов при изменении порогового значения результирующей функции	82
4.2	Интерфейс системы идентификации минералов на базе интерактивного реестра	85
4.3	Распределение значений результирующей функции и их интерпретация в соответствии с выбранным пороговым значением	86

Список таблиц

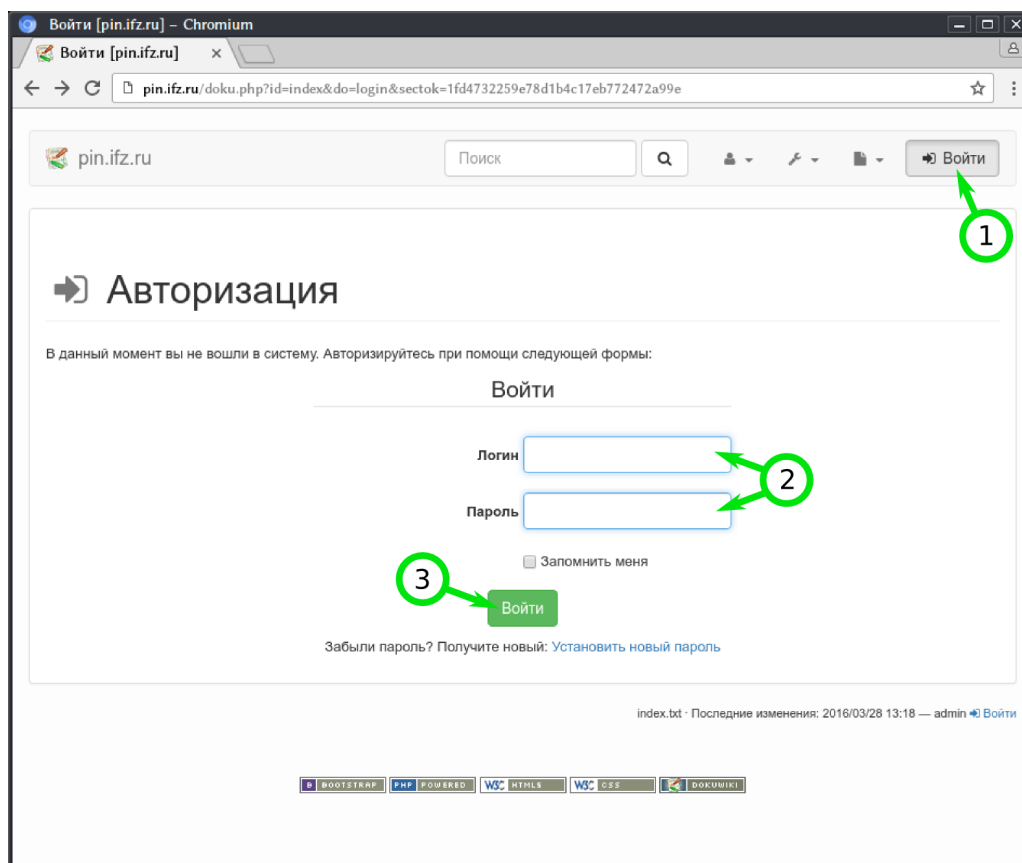
2.1	Сравнение CMS MediaWiki и DokuWiki	35
2.2	Сравнение некоторых синтаксических элементов языков C, R и Python	46
2.3	Примеры нотации названий столбцов	49
2.4	Структура объекта, возвращаемого методами построения дискриминационных диаграмм	60

Список листингов

2.1	Шаблон страницы метода в интерактивном реестре	38
2.2	Шаблон страницы библиографической информации в интерактивном реестре	39
2.3	Шаблон исходного кода геотермометра	48
3.1	Пример описания сенсора в программе TPF	71
4.1	Ограничения, налагаемые на состав биотита	79
4.2	Вывод исполняемого модуля идентификации	83
4.3	Набор правил для определения фенгита	87

Приложение 1. Основные операции в интерфейсе интерактивного реестра

Вход в систему



Интерфейс входа в реестр / систему DokuWiki:

1. Кнопка, открывающая диалог входа;
2. Поля для ввода регистрационных данных;
3. Кнопка входа в систему.

Главная страница реестра

The screenshot shows the main page of the registry (pin.ifz.ru) in a Chromium browser. The page features a navigation bar with a search box and a main content area with several sections. Five green circles with numbers 1 through 5 are placed over specific links:

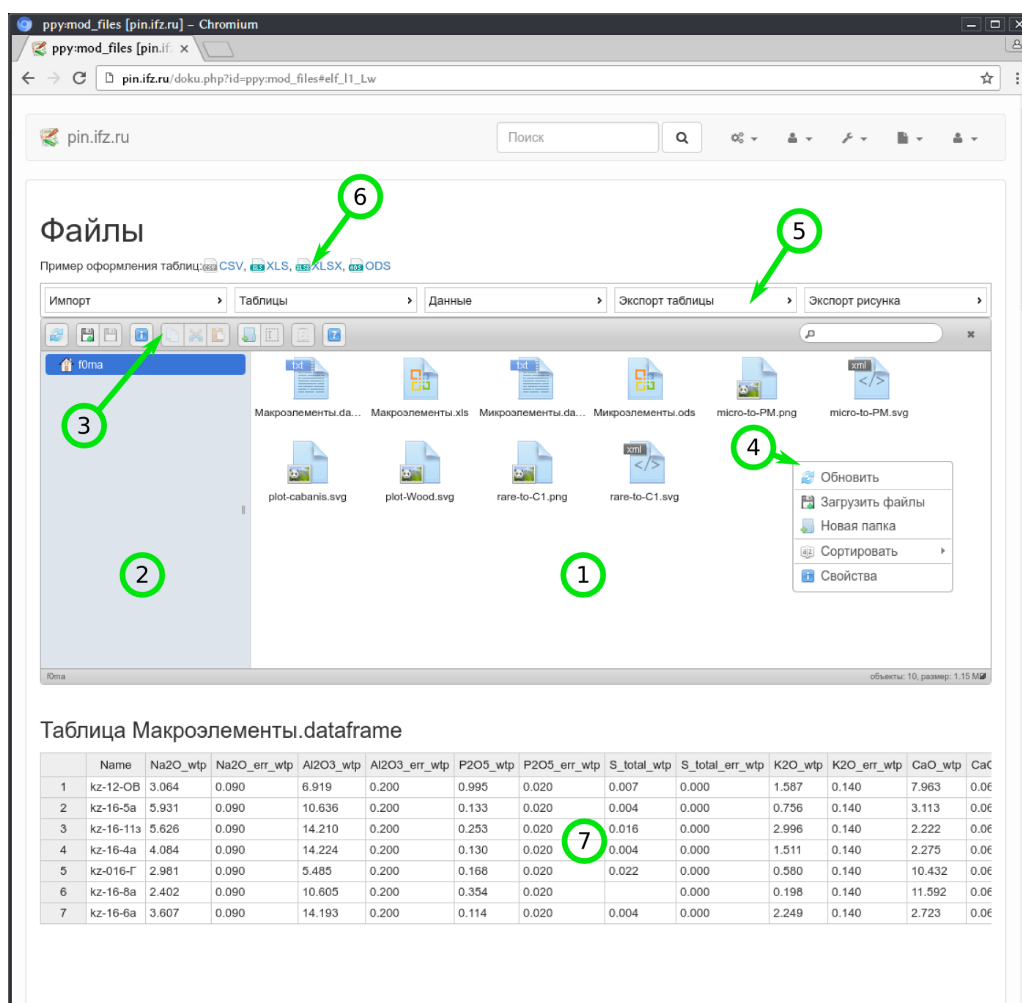
- 1: Points to the link "Менеджер файлов" in the "Управление данными" section.
- 2: Points to the link "Базовая визуализация" in the "Управление данными" section.
- 3: Points to the link "Интерпретация состава минерала и разложение на минералы" in the "Петрологические инструменты" table.
- 4: Points to the link "Методы интерактивной вики" in the "Петрологические инструменты" table.
- 5: Points to the "Рекомендации" link in the "Источники" section.

The page also includes a notification banner at the top, a table of petrological tools, and various navigation links at the bottom.

Главная страница реестра / главная страница DokuWiki:

1. Ссылка на менеджер файлов;
2. Ссылка на модуль базовой визуализации;
3. Ссылки на модули интерактивного реестра;
4. Ссылки на соответствующие модулям пространства имен вики;
5. Дополнительные ссылки.

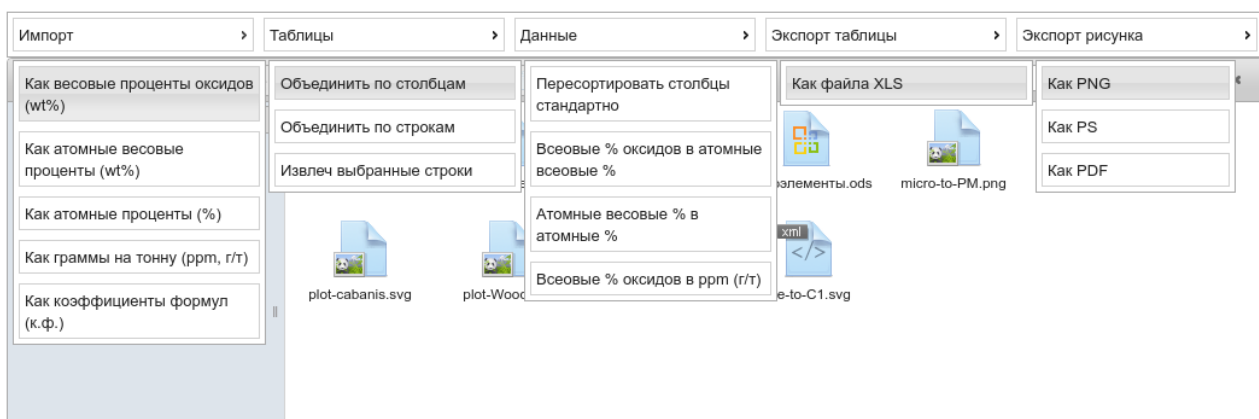
Управление файлами пользователя



Интерфейс управления файлами:

1. Обзор файлов в текущей папке (файлы таблиц, наборов данных (.dataframe) и изображений (.svg и .png));
2. Дерево папок пользователя;
3. Панель операций файлового менеджера eFinder: обновить, загрузить, скачать, просмотр сведений о файле, копировать, вырезать, вставить, создать папку, переименовать файл, удалить, о программе eFinder, поиск, очистка поиска;
4. Контекстное меню файлового менеджера;
5. Меню операций с файлами (см. ниже);
6. Образцы файлов в различных форматах (CSV, XLS, XLSX, ODS);
7. Автоматический просмотр таблиц в формате .dataframe.

Меню операций с файлами



Все меню операций с файлами, открытые одновременно:

1. Меню Импорт: преобразование файлов из распространенных табличных форматов (CSV, XLS, XLSX, ODS) в формат `.dataframe`, используемый в реестре. Преобразованные файлы становятся доступны в модулях. Выбор варианта импорта определяет, каким образом будут преобразовываться заголовки столбцов в соответствии с правилами, рассмотренными в подразделе 2.3.4;
2. Меню Таблицы: базовые операции с таблицами. Объединение по строкам, по столбцам и извлечение выбранных строк в отдельный файл. Обрабатываются только файлы `.dataframe`;
3. Меню Данные: сортировка столбцов таблицы и преобразование форматов данных (весовые проценты оксидов в атомными весовыми проценты, и ppm).
4. Меню Экспорт таблицы: преобразование файла `.dataframe` в распространенный формат электронных таблиц XLS;
5. Меню Экспорт рисунка: преобразование рисунка, созданного реестром из формата `.svg` в один из распространенных графических форматов.

Корневая страница пространства имен

classifydiag [pin.ifz.ru] – Chromium
classifydiag [pin.ifz.ru] x
pin.ifz.ru/doku.php?id=classifydiag

Методы построения классификационных диаграмм 1

Содержание

- Методы построения классификационных диаграмм

For whole rock

- Alk-MgO-FeO: Tholeiitic - Calc-Alkaline
- FeO/MgO-SiO₂: Tholeiitic-Calc-alkaline basalts
- K₂O-SiO₂: Basic-Acid basalts
- La-Y-Nb: Continental crust magmatic mixtures, Cabanis, 1989
- MgO-FeO-Al₂O₃: Crust type, Pearce 2
- MnO-TiO₂-P₂O₅: Petrotectonic Environments, Mullen, 1983
- Nb-Zr-Y: Geodynamic Settings, Meschede
- TAS Classification
- Th-Hf-Nb: Wood
- Y-Cr: Petrotectonic Environments, Pearce, 1982
- Zr-Ti-Y: Geodynamic Settings, Pearce
- Диаграмма классификации горных пород

For minerals

- Px: Nomenclature of pyroxenes in Ca-Mg-Fe, Morimoto, 1988
- Px: Q-J Diagram, Morimoto, 1988
- True and brittle micas composition, Tischendorf, 2007

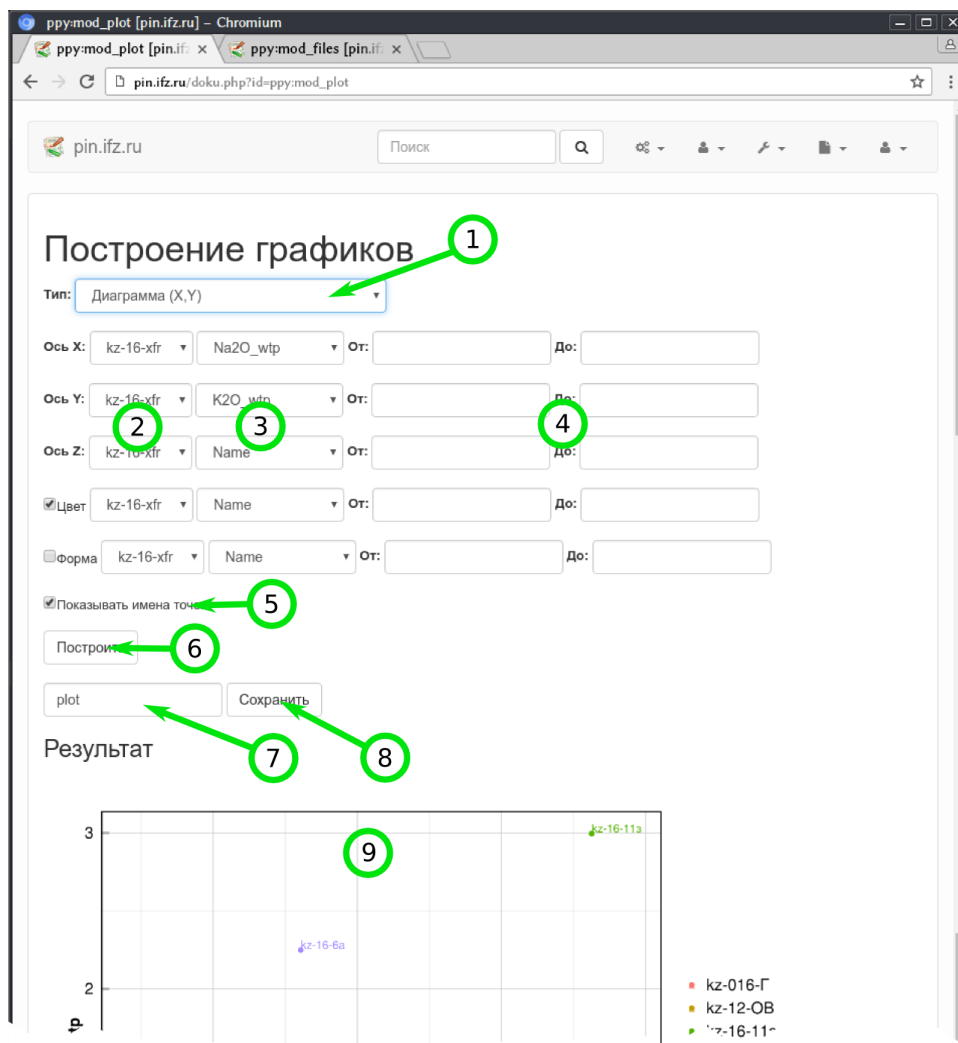
All pages in category

A	M (продолжение)	T (продолжение)
<ul style="list-style-type: none">Alk-MgO-FeO: Tholeiitic - Calc-Alkaline	<ul style="list-style-type: none">MnO-TiO₂-P₂O₅: Petrotectonic Environments, Mullen, 1983 3	<ul style="list-style-type: none">Th-Hf-Nb: Wood
<ul style="list-style-type: none">FeO/MgO-SiO₂: Tholeiitic-Calc-alkaline basalts	<ul style="list-style-type: none">Nb-Zr-Y: Geodynamic Settings, Meschede	<ul style="list-style-type: none">True and brittle micas composition, Tischendorf, 2007
<ul style="list-style-type: none">K		<ul style="list-style-type: none">Y-Cr: Petrotectonic Environments,

Области на корневой странице пространства имен:

1. Заголовок пространства;
2. Список страниц / методов по категориям;
3. Общий список страниц / методов в пространстве имен.

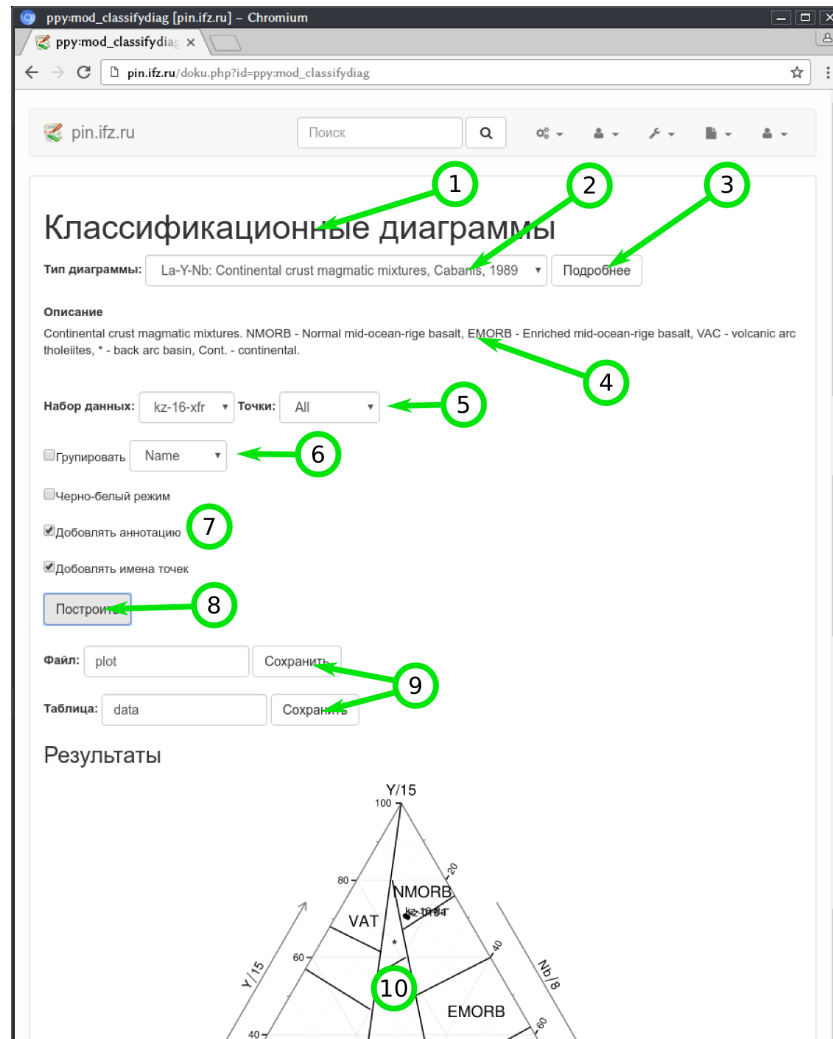
Базовая визуализация



Интерфейс базовой визуализации:

1. Выбор типа графика: гистограмма, диаграмма, тригонограмма, диаграмма Эндрюса;
2. Выбор набора данных для построения;
3. Выбор столбца из набора данных. Набор и столбец определяется для каждой оси отдельно. Поддерживаются оси X, Y, Z, цвет и форма маркеров;
4. Диапазоны значений для каждой оси;
5. Показывать на графике имена точек;
6. Сохранение файла: имя файла;
7. Сохранение файла: кнопка сохранения;
8. Просмотр полученного изображения.

Модуль специальной визуализации



Интерфейс модуля специальной визуализации:

1. Имя модуля;
2. Выбор метода визуализации;
3. Ссылка на страницу метода;
4. Выбор набора данных и отдельных строк из него;
5. Опция группировки по заданному столбцу;
6. Опции построения графика;
7. Кнопка построения;
8. Сохранение файла изображения: поле ввода имени и кнопка сохранения, сохранение таблицы промежуточных результатов: поле ввода имени и кнопка сохранения;
9. Просмотр полученного изображения.

Модуль геобаротермометрии

pin.ifz.ru

Оценка PT условий методами геотермобарометрии

Тип: Temperature sensors

Сенсор: Bt: Henry, Guidotti, Thomson, 2005

Описание

Исходные данные

В: Bt

265-2_4 т.3

Диапазоны

Использовать рассчитанные коэффициенты формул

Рассчитать

Файл: calc

Сохранить

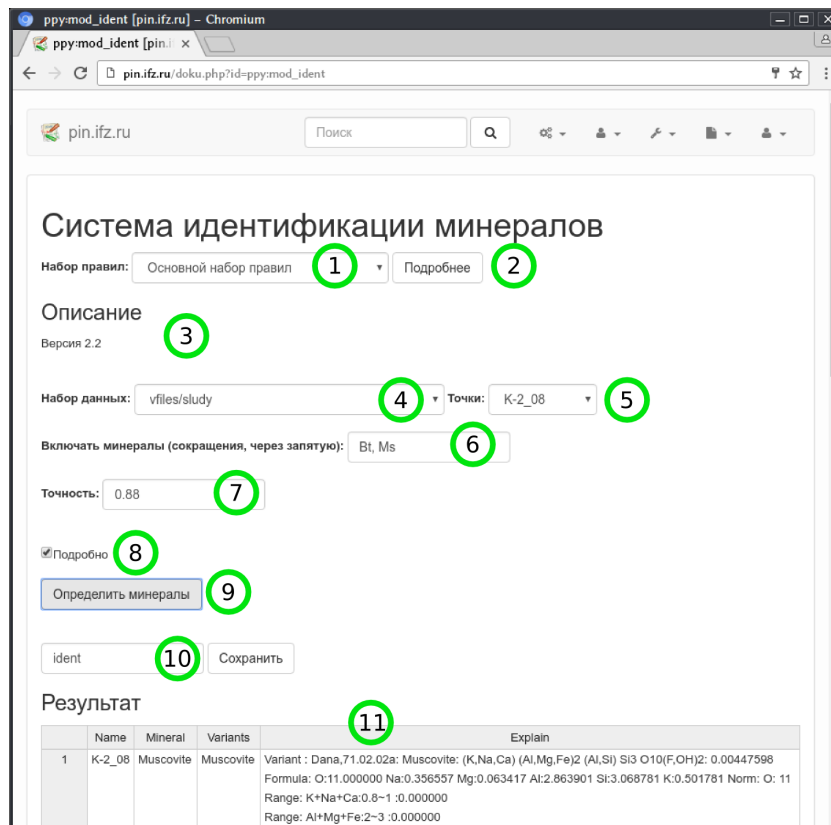
Результат

	Name	
1	265-2_4 т.3	741.042

Интерфейс модуля геобаротермометрии:

1. Выбор группа сенсоров;
2. Выбор конкретного сенсора;
3. Переход на страницу метода;
4. Краткая информация о методе;
5. Выбор исходного набора данных (файла .dataframe);
6. Выбор конкретного анализа;
7. Использовать в качестве входных данных коэффициенты формул;
8. Кнопка выполнения расчета;
9. Поле для указания имени файла и кнопка сохранения результатов;
10. Таблица результатов.

Модуль идентификации минералов



Интерфейс модуля идентификации минералов:

1. Выбор набора ограничений;
2. Переход на страницу набора ограничений;
3. Краткая информация о наборе;
4. Выбор исходного набора данных (файла `.dataframe`);
5. Выбор конкретного анализа;
6. Список минералов, которые будут включены в отчет, даже если значение результирующей функции для них окажется меньше порогового (см. подраздел 4.1);
7. Пороговое значение результирующей функции (см. подраздел 4.1);
8. Выводить подробный отчет при определении;
9. Кнопка выполнения определения;
10. Поле для указания имени файла и кнопка сохранения результатов;
11. Таблица результатов.