

Отзыв на диссертацию С.Д. Иванова

«ПРИМЕНЕНИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ГЕОБАРОТЕРМОМЕТРИИ И СМЕЖНЫХ ЗАДАЧАХ»,

представленной на соискание степени кандидата технических наук по специальности 25.00.10 – «Геофизика, геофизические методы поисков полезных ископаемых»

Поиск оптимальных способов хранения и систематизации как разнородных данных, так и методов их обработки с предоставлением к ним широкого доступа стал насущной задачей в области применения информационных технологий, определяющей одно из направлений их развития. Представленная работа посвящена разработке одного из решений этой несомненно актуальной проблемы в приложении к минералого-петролого-геохимическим данным и методам.

Эта работа является по сути междисциплинарной, поэтому будет логичным разделить её на две составляющие. Первая из них относится непосредственно к информационным технологиям (ИТ). Автор проанализировал некоторые из имеющихся в настоящее время популярных разработок из рассматриваемой области применения ИТ, выявил их главные недостатки и составил ряд обоснованных требований, которым должна соответствовать современная система для хранения и обработки результатов научных исследований. Затем на основе анализа существующих в настоящее время ИТ-решений автор предложил архитектуру веб-приложения, построенную на технологии вики в связке с разработанными им модулями для хранения и управления как пользовательскими данными, так и методов их обработки. Созданный автором действующий программный продукт, на мой взгляд, достаточно убедительно показывает, что предложенное решение жизнеспособно – и, вместе с тем, лишено тех недостатков, которые были выявлены у предшественников.

Вместе с тем, стоит заметить, что недостатки существующих инструментов не являются непреодолимыми, хотя и существенно затрудняют их использование, бесспорно снижая эффективность научной работы.

Например, полнота справочной информации зависит в большей мере от наличия времени и трудолюбия авторов библиотек методов и, как правило, не является ограничением, присущим непосредственно приложению, его архитектуре. Это положение, в частности, демонстрируется и на примере разработанного автором интерактивного реестра, в котором значительная часть реализованных геотермобарометрических методов пока что остаётся недокументированной. С другой стороны, зачастую сделать компетентный вывод о применимости того или иного метода можно, лишь ознакомившись с самой статьёй-первоисточником (а также со статьями, ссылающимися на неё и предлагающими модификации метода, лишённые выявленных недостатков). То-есть, именно квалификация исследователя, выбирающего метод расчёта, является основным определяющим фактором.

Закрытость исходных кодов также больше зависит от авторов программных продуктов: многие из них предоставляются вместе с исходными текстами – либо полностью, либо лишь в отношении реализаций самих научных методов (с закрытыми интерфейсным и др. служебными разделами). Как ни странно, эта же проблема обнаружилась и при ознакомлении с интерактивным реестром, в котором исходные коды для двух (из четырёх) наборов правил идентификации минералов недоступны.

Однако, стоит отметить ещё одну серьёзную проблему, которую автор преодолевает своей разработкой: это недостаточный уровень унификации существующих реализаций методов, который не позволяет достаточно быстро объединить их в единой базе данных. В частности, использование различных языков

программирования, соглашений вызовов, структур данных (вплоть до разной архитектуры приложений) сильно затрудняют анализ ключевых алгоритмов методов для их использования в сторонних продуктах.

Проблема платформозависимости в настоящее время достаточно успешно решается как с помощью виртуальных машин, так и просто использованием необходимого аппаратно-программного обеспечения.

Несмотря на свою привлекательность и несомненную перспективность, существующие веб-технологии также обладают рядом недостатков, главные из которых: проблемы с доступностью и устойчивостью работы сервисов, а также относительно невысокая скорость работы приложений. Эти недостатки связаны с высокой сложностью систем и их зависимостью от очень большого числа внешних факторов (в т.ч. от качества интернет-соединения и конфигураций локальных систем), а также с использованием языков-интерпретаторов для реализаций методов (в частности, R). Последнее может быть не столь заметно при обработке совсем небольших массивов данных, но становится критичным при увеличении их объёмов и сложности методов. Для преодоления одного перечисленных недостатков – недоступности веб-сервиса, – автором предусмотрена выгрузка методов на локальный компьютер пользователя.

Вторая составляющая работы относится к области минералогии и петрологии (идентификация минералов и геотермобарометрия), а также к геохимии (дискриминационные диаграммы и спайдерграммы). В связи с этим возникает некоторое сомнение в правильности выбора специальности для диссертационной работы («Геофизика, геофизические методы поисков полезных ископаемых»): в её паспорте ВАК я просто не нашёл пунктов, отвечающих содержанию работы.

Для наполнения геотермобарометрического раздела интерактивного реестра автором был успешно проведён импорт наиболее крупной в настоящее время библиотеки, содержащей формализованные алгоритмы методов – базы сенсоров программы TRF. Кроме того, в реестре реализованы пересчёты химических анализов минералов на их кристаллохимические формулы (с вычислением дополнительных коэффициентов, требуемых как для геотермобарометрии, так и для построения классификационных диаграмм). Тестирование (проведённое как автором, так и независимо рецензентом) показывает, что предложенные решения являются вполне работоспособными и имеют перспективы для дальнейшего развития. В частности, к пожеланиям можно отнести автоматическое сокращение набора предлагаемых методов на основе простого анализа минеральных составов изучаемых образцов, а также визуализацию результатов геотермобарометрического исследования в виде диаграмм в координатах P , T , fO_2 , aH_2O и т.д.

Весьма интересным и важным является раздел, посвящённый методике автоматической идентификации минералов на основе их химических составов – и к которому относится третье защищаемое положение. По мнению рецензента, оно явно недоработано и положением не является, т.к. представляет собой сложноподчинённое предложение без сказуемого в главной части, лишь осложнённое причастным оборотом и придаточным определительным. Возможный вариант решения – удаление союзного слова "которая": это превратило бы глагол "решает" в сказуемое главной части и сформировало положение "система решает", которое и следует защищать.

Задачу идентификации минералов автор предлагает решать с помощью аппарата нечёткой логики. У меня есть только самые общие представления об этом подходе, поэтому следующие вопрос и рассуждение можно частично считать и плодом моей некомпетентности: а можно ли вообще, применяя этот метод, проводить достаточно глубокий анализ тех причин, по которым те или иные анализы определяются им неправильно?

Если можно, то насколько это просто, какие затруднения при этом возникают и есть ли опыт повышения достоверности идентификации после подобного анализа и соответствующей коррекции правил? Можно ли этот анализ формализовать (программно реализовать) и предоставить его результаты пользователям системы?

Если анализ невозможен или слишком затруднителен, то возникает некоторое сомнение в большой научной ценности инструмента, который является тогда по сути искусственным если не «чёрным», то «серым ящиком» (т.к. процессы в нём анализу с большим трудом, но всё же поддаются), даже несмотря на его успешное чисто утилитарное применение. По-моему, если эксперт (человек или заменяющая его компьютерная система) начинает колебаться между несколькими вариантами и в конце-концов приходит к некоторому решению, то он должен предоставлять некоторые внятные аргументы в его пользу, не полагаясь лишь на «научную интуицию» (эквивалентом которой можно считать абстрактные комплексные коэффициенты, вычисляемые экспертной системой). При отсутствии таких аргументов (либо слишком высоком уровне их абстрактности) решение неизбежно теряет свою научную ценность. Эта проблема, кстати, присуща многим работам, посвящённым использованию подобных алгоритмов в научных исследованиях.

Стоит также упомянуть и о наиболее существенном негативном свойстве экспертных систем определения минералов по химическим составам, в которых базы данных основаны на подобных описаниях через диапазоны содержаний элементов (лишь за исключением правила сравнения "more than" – кстати, реализованного в работающей вики-системе, но не описанного в тексте диссертации). Тут стоит упомянуть, что такие экспертные системы уже существуют: например, на подобных описаниях построена идентификационная часть микроаналитического аппаратно-программного комплекса QEMSCAN (Haberlah et al., 2011), хотя собственно механизм принятия решений там, вероятно, построен на других принципах. Проблема состоит в том, что описания, основанные только на диапазонах содержаний элементов и не включающие ограничения по балансу зарядов, определяют слишком обширные пространства составов минералов-твёрдых растворов (хоть они и создают видимость сохранения стехиометрических соотношений). Значительная доля изоморфных замещений в минералах является гетеровалентными, что существенно сокращает пространства их допустимых составов за счёт строгого согласования содержаний элементов с разными степенями окисления. При достаточно большом числе минералов, обладающих сходными составами, это неизбежно должно приводить к росту числа некорректных и/или неоднозначных определений – просто в силу того, что увеличенные (некорректные) пространства для некоторой пары минералов могут оказываться пересекающимися, тогда как их же пространства, но с исключёнными несбалансированными по зарядам частями, могут быть непересекающимися (т.е. в принципе допускать однозначное определение).

Лично я пока что не уверен в том, что в экспертной системе, разрабатываемой автором диссертации, можно использовать ограничения на баланс зарядов при одновременном сохранении корректных стехиометрических соотношений.

Возможно, этот недостаток и не будет играть столь большой роли при весьма небольшом объёме базы данных по составам минералов, который требуется для геотермобарометрических исследований. Однако, при попытках её существенного расширения и превращения в некий универсальный инструмент для идентификации минералов (а это, по-видимому, должно произойти как с ростом коллектива, участвующего в разработке вики-системы, так и просто по прошествии некоторого времени) он должен стать заметной проблемой. Вместе с тем, надо отметить, что и более продвинутых веб-систем пока что также не разработано.

Для апробации методики автором был взят массив из почти полутора тысяч анализов минералов, который был обработан с использованием заданных вручную правил для идентификации более, чем трёх десятков минералов. В результате было получено примерно 91% верных определений, т.е. примерно одно неверное определение на десять верных. По мнению рецензента, это соотношение трудно назвать вполне удовлетворительным – конечно же, при условии достаточно качественных аналитических данных и отсутствии действительных пересечений химических составов минералов, которые также могут присутствовать (например, у некоторых амфиболов с пижонитами). Тут возникает очевидное желание посмотреть: какие именно анализы попали в группу неверно определённых? Однако, этот важный и необходимый фактический материал (непосредственные результаты численных экспериментов и детальный анализ причин неверных определений) в диссертационной работе не приводится, в том числе и среди приложений в самом её конце.

Независимое тестирование веб-системы, проведённое рецензентом, продемонстрировало результаты, которые следует считать скорее неудовлетворительными: в небольшом массиве из 16 вполне качественных анализов 14-ти минералов, имеющих в основном набор правил (Ap, Vt, Cpx, Grt, Hbl, Ilm, Kfs, Mus, Ol, Opx, Pl, Spl, Srp, Zrc), были верно определены только 5 из них, тогда как остальные либо неверно, либо остались неопределёнными. Правда, я не могу исключить, что это поведение связано с посторонними причинами: например, тестовым изменением автором правил, рабочей модификацией алгоритмов или настроек.

По мнению рецензента, одной из весьма насущных проблем, ставшей особо острой с развитием крупных справочных информационных систем, является появление больших количеств ошибок и опечаток, неизбежно возникающих при их ручном наполнении недостаточно квалифицированным и/или невнимательным персоналом. К сожалению, это можно продемонстрировать и на примере данной работы, включив в рассмотрение и сам веб-ресурс. Скажем, в описании задачи идентификации минералов в качестве примера взят биотит, для которого в базе данных приводилась некорректная формула с двумя атомами кислорода и 211 атомами кремния, а в гидроксильной позиции вместо ОН записана H₂O. Хотя эти формулы являются справочными и не участвуют непосредственно в расчётах, они могут вводить в заблуждение и, что наиболее критично – являться эпицентрами распространения ошибок, что весьма характерно для информационных веб-систем с простым и широким доступом типа вики. Эти опечатки, присутствующие в тексте автореферата, в настоящее время на веб-ресурсе исправлены. Но в вики-системе в настоящее время (на момент написания рецензии) обнаруживаются другие недоработки: например, для магнезита приводится формула, скорее отвечающая магнетиту (Mn,Fe,Mg)(Fe,Mn)2O₄, в его наборе правил нормализация производится на 12 кислорода (впрочем, как и для большинства остальных минералов основного набора) и обязательным элементом указан Mn; для рутила записано единственное правило Ti,Nb,Fe,Ta,Mn,Sn,Pb,Ge=6 (при его пересчёте опять таки на 12 кислорода); плагиоклаз может содержать существенные количества Fe и Mg. Не лишены опечаток и другие разделы ресурса: например, на небольшой по объёму тексте странице «Константы» мы встречаем «Continental Crust composition» (вместо «Crust»), «Carbonaceous Chondrites» (вместо «Carbonaceous»), «Бронеман-Старынкевич» (вместо «Борнеман-Старынкевич»).

Поэтому одной из необходимых задач оказывается также разработка программных средств проверки на ошибки ввода по крайней мере для тех объектов, которые допускают подобную автоматическую проверку. Формулы минералов и правила идентификации как раз относятся к таким объектам. В связи с этим возникает и такой вопрос: предусмотрена ли в разработанной системе идентификации минералов проверка вводимых вручную правил, чтобы они не оказывались

взаимоисключающими? Однако, в любом случае, один из самых важных принципов при ручном наполнении различных баз данных и прочих информационных ресурсов – это предельная внимательность при работе, когда работник изначально старается не допускать опечаток.

По мнению рецензента, работа соответствует требованиям, установленным п. 9 Положения о присуждении учёных степеней (утверждено постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842) для ученой степени кандидата наук, а её автор Иванов Станислав Дмитриевич заслуживает ученой степени кандидата технических наук.

Рецензент:

Старший научный сотрудник
Федерального государственного
бюджетного учреждения науки
Института геологии и
геохронологии докембрия
Российской академии наук
(ИГГД РАН), кандидат геолого-
минералогических наук

Дмитрий Владимирович Доливо-Добровольский

10.04.2017

199034, Санкт-Петербург,
Наб. Макарова, д. 2. ИГГД РАН
Тел. +7(921)554-98-73 (моб.), 8(812)328-03-62 (сл.)
E-mail: ddd@ipgg.ru, doliva@inbox.ru

Автор отзыва дает согласие на включение и обработку своих персональных данных для целей, связанных с работой диссертационного совета.

