

УДК 81.32

ВЫБОР ОПТИМАЛЬНОГО РАДИУСА КЛАСТЕРИЗАЦИИ НА КАРТАХ КАТАСТРОФ

А. А. Крижановский

*Институт прикладных математических исследований КарНЦ РАН,
ФИЦ «Карельский научный центр РАН» (ул. Пушкинская, 11, Петрозаводск,
Республика Карелия, Россия, 185910)*

Викиданные содержат информацию о вулканах и землетрясениях, позволяющую провести их пространственный анализ. Чтобы найти взаимосвязь между ними, нужно их сгруппировать в кластеры. Рассмотрено несколько подходов выбора оптимального радиуса кластеризации. Варьирование радиуса кластеризации показало, что (1) 90 км – оптимальное расстояние для выявления парных взаимодействий (максимум смешанных кластеров, содержащих и вулканы, и землетрясения); (2) 127 км – порог устойчивости кластерной структуры (превышение ведет к артефактам слияния). С помощью SPARQL-запросов и программ на языке Python были построены карты, отображающие возможные взаимосвязи между извержениями и сейсмическими событиями. Анализ показывает, что сейчас в Викиданных мало информации о датах извержения вулканов, хотя число объектов «извержение вулкана» со свойством «дата» и выросло за три года в несколько раз. Показаны возможности и текущие ограничения Викиданных для геонаучных исследований. Работа подтверждает, что краудсорсинговые базы данных могут дополнять традиционные научные ресурсы при исследовании глобальных геофизических закономерностей.

Ключевые слова: землетрясение; извержение вулкана; кластеризация; Викиданные

Для цитирования: Крижановский А. А. Выбор оптимального радиуса кластеризации на картах катастроф // Труды Карельского научного центра РАН. 2025. № 4. С. 24–33. doi: 10.17076/mat2099

A. A. Krizhanovsky. SELECTING THE OPTIMAL CLUSTERING RADIUS ON DISASTER MAPS

Institute of Applied Mathematical Research, Karelian Research Centre, Russian Academy of Sciences (11 Pushkinskaya St., 185910 Petrozavodsk, Karelia, Russia)

Wikidata contains data on volcanoes and earthquakes, enabling spatial analysis. To find the relationship between them, they need to be grouped into clusters. Several approaches to choosing the optimal clustering radius are considered. Varying the clustering radius revealed that (1) 90 km is the optimal distance for detecting pairwise interactions (maximizing mixed clusters containing both volcanoes and earthquakes), and (2) 127 km is the stability threshold of the cluster structure (exceeding it leads to merging artifacts). Using SPARQL queries and Python scripts, maps were generated to visualize potential connections

between eruptions and seismic events. The analysis shows that Wikidata currently lacks sufficient information on volcanic eruption dates, although the number of “volcanic eruption” objects with date properties has increased severalfold over three years. The study highlights both the potential and current limitations of Wikidata for geoscientific research. The study confirms that crowdsourced databases can complement traditional scientific resources in researching global geophysical patterns.

Key words: earthquake; volcanic eruption; clustering; Wikidata; SPARQL

For citation: Krizhanovsky A. A. Selecting the optimal clustering radius on disaster maps. *Trudy Karelskogo nauchnogo tsentra RAN = Transactions of the Karelian Research Centre RAS*. 2025. No. 4. P. 24–33. doi: 10.17076/mat2099

ВВЕДЕНИЕ

Вулканы и землетрясения с древних времен привлекали внимание людей и фиксировались учеными. Одной из крупнейших современных баз данных, описывающих конкретные извержения вулканов и землетрясения, является Викиданные. В работе поставлена двоякая задача: найти пространственную взаимосвязь вулканической и сейсмической активности на основе информации Викиданных и оценить степень полноты самой этой базы.

Мы не претендуем на новые открытия в области вулканологии и сейсмологии. Но хотим взять ту часть базы Викиданных, которая описывает вулканы и землетрясения, и оценить полноту информации и возможность проведения экспериментальных исследований на основе этой огромной и доступной базы данных. То есть мы хотим проверить, насколько Викиданные готовы, чтобы на их основе проводить исследования. Вулканы и землетрясения – это два типа объектов из тысяч других в Викиданных. Положительный ответ позволит с большей уверенностью исследовать и другие стороны Вселенной на основе Викиданных.

В следующем разделе представим обзор литературы по вопросу взаимосвязи вулканов и землетрясений. Далее опишем доступные базы данных вулканов, в том числе Викиданные. Затем извлечем эти сведения из Викиданных с помощью SPARQL-запросов [1]. Остановимся на вопросе выбора оптимального порога при кластеризации пространственных данных. Затем мы опишем алгоритм программы поиска и визуализации на карте ближайших друг к другу вулканов и землетрясений и обсудим недостатки и потенциал Викиданных.

ОБЗОР ИССЛЕДОВАНИЙ, ИЗУЧАЮЩИХ ВЗАИМОСВЯЗЬ ЗЕМЛЕТРЯСЕНИЙ И ВУЛКАНОВ

В работе [14] проверили гипотезу, что вулканы и землетрясения связаны во времени

и пространстве. Гипотеза подтвердилась статистически, были найдены последовательности «извержение – землетрясения» [14]. Найдена связь между крупными извержениями и землетрясениями во времени и пространстве на примере исторических записей активности вулкана Мауна-Лоа, второго по объему и площади подошвы вулкана на Земле.

Часто отмечают, что крупные, не слишком отдаленные землетрясения влияют на извержения вулканов. В работе [2] проверялась гипотеза: следуют ли за землетрясениями извержения вулканов. Для индонезийского вулканического региона в целом получен отрицательный ответ. Однако если еще дополнительно учитывать время извержения, расстояние и силу землетрясения, то ответ будет иным, а именно: влияние землетрясения было статистически значимым для 7 из 35 вулканов, но только после учета внутреннего состояния вулкана [2]. Тектонические землетрясения (то есть вызванные смещением горных плит) могут приводить в действие (вызывать) большинство типов вулканов, но только при выполнении ряда вулканических и сейсмических условий [11].

Есть трудности в учете и анализе извержений вулканов:

- *Извержение имеет длительность, а не точку во времени.* «Извержение может длиться очень долгое время, например, вулкан Стромболи действует уже более 2500 лет» [13, с. 435].
- *Рост частоты исторических наблюдений.* Число активных вулканов в год стремительно растет последние пять веков. Но это объясняется ростом населения и лучшей коммуникацией, растет (читательский) интерес к вулканам. Это объясняет снижение роста числа вулканов во время двух мировых войн [13, с. 436].

- *Цикличность и эпизодичность.* Ряд вулканов, наблюдаемых сотни лет (например, Везувий, Колима, Майон), показывают повторяемость, цикличность извержений. С другой стороны, вулканологи при ограниченных ресурсах должны максимум внимания уделять именно опасным вулканам, но их выбор только на основе записей о предыдущих извержениях будет ошибкой, поскольку за последние 200 лет каждый год происходит в среднем одно или два извержения вулканов, отсутствующих в каких-либо записях [13, с. 446]. Другие вулканы извергаются после сотен лет тишины.

Внешние явления (землетрясения, оползни и приливы) могут потревожить магматический резервуар и вызвать извержение вулкана [3]. В среднем в год извергается 50 вулканов.

Извержения вулканов чаще происходят после землетрясений [6]. «Исторически сложилось так, что некоторые вулканы извергались после крупных землетрясений. . . Статистический анализ мировых вулканических и сейсмических записей показывает значительное увеличение вероятности извержения после землетрясения в течение периода от нескольких дней до нескольких лет» [3].

БАЗЫ ДАННЫХ ВУЛКАНОВ

Проблемы баз данных вулканов в том, что нет единого стандарта, мало доступных баз данных [4, с. 739].

База вулканов всего мира WOVodat (<https://wovodat.org/>) служит ресурсом для улучшения прогнозирования извержений. Сходство между Викиданными и WOVodat в том, что «база данных связана ссылками с различными открытыми базами и каталогами, обеспечивающими многообразие дополнительных данных» [4, с. 739].

Наше исследование требует наличия обширной базы данных по вулканам и землетрясениям. Однако попытки создать глобальные вулканические базы данных начались лишь недавно [9]. Автор статьи [9] предлагает объединить две открытые базы данных о вулканах – GVP и LaMEVE – в единую. Это база данных извержений голоцена проекта Global Volcanism Project (GVP) Смитсоновского института (<http://volcano.si.edu>) и база данных мощных взрывных вулканических извержений LaMEVE [5], созданная в рамках проекта VOGRIPA (<http://www.bgs.ac.uk/vogrifa>). Таким образом, в объединенной базе GVP+LaMEVE содержится информация о 9517 извержениях с указанием показа-

теля вулканической эксплозивности (VEI) [9]. Сравните: 10 415 извержений вулканов в хронологическом порядке приводятся в книге [12].

Викиданные развиваются как связующий центр информации [15, с. 868], описанные выше две базы вулканов частично включены в Викиданные.

Видим на рис. 1, что объекты Викиданных, описывающие вулканы, могут содержать ссылки на те же вулканы в базах GVP и VOGRIPA. Ссылки на эти базы единообразно оформляются с помощью внешних идентификаторов Викиданных, а именно: Global Volcanism Program ID (P1886), VOGRIPA ID (P4708). Таким образом, Викиданные в некоторой степени объединяют эти базы данных вулканов и извержений.

Mount Tandikat (Q40692)...

Item [Discussion](#)

stratovolcano in West Sumatra, Indonesia
Gunung Tandikat | Tandikat

[In more languages](#)

Language	Label	Description
English	Mount Tandikat	stratovolcano in West Sumatra, Indonesia

Identifiers

Global Volcanism Program ID	261150 ...	edit
	0 references	
VOGRIPA ID	811 ...	edit
	0 references	+ add referen
		+ add value

Рис. 1. Фрагмент карточки вулкана Mount Tandikat (Q40692) в Викиданных с идентификаторами *Global Volcanism Program ID* и *VOGRIPA ID*
Fig. 1. A fragment of the Mount Tandikat volcano card (Q40692) in Wikidata with the identifiers *Global Volcanism Program ID* and *VOGRIPA ID*

SPARQL-ЗАПРОСЫ ПО ИЗВЛЕЧЕНИЮ ДАННЫХ О ВУЛКАНАХ И ЗЕМЛЕТРЯСЕНИЯХ: ПОДГОТОВКА ДАННЫХ

Викиданные пока не так полны, как нам хотелось бы, а именно: почти нет данных о времени и дате извержений вулканов, нет данных о внутреннем состоянии вулканов. Посмотрим, какую ценную информацию о вулканах и землетрясениях мы можем получить из Викиданных.

Обычно приводят листинги со SPARQL-запросами. Для экономии места вместо семи листингов представим данные в таблице. Чтобы увидеть листинг любого из этих запросов, щелкните по ссылке для перехода в Wikidata Query Service [1]. Откроется окно в браузере с кодом запроса. Далее щелкните по кнопке *Play*, чтобы выполнить запрос. После выполнения запроса на той же странице ниже будет результат: список, график или карта.

Наше исследование длится несколько лет, поэтому у нас есть данные за три года (2023–

2025). Таблица показывает, что Викиданные постоянно меняются. Выросло число и вулканов, и землетрясений. Важным для исследования является рост числа вулканов и землетрясений, имеющих координаты (запросы 6 и 7 в таблице).

Было бы интересно узнать, в какой пропорции добавляются объекты, появившиеся за эти три года (например, произошли извержения), и в какой доле добавляются «старые» вулканы и землетрясения, проявившие себя до 2023 года.

SPARQL-запросы и результаты по вулканам и землетрясениям в Викиданных
SPARQL queries and the results for volcanoes and earthquakes in Wikidata

№ No.	Задача Task	Количество / Number			Ссылка Link
		2023	2024	2025	
1	Список всех землетрясений List of all earthquakes	2394	2430 ↑1.5 %	2451 ↑0.9 %	w.wiki/A362
2	Список всех вулканов List of all volcanoes	1891	–	2052	w.wiki/Dyad
3	Количество землетрясений по странам (график). Число стран с землетрясениями Number of earthquakes by country (chart). Number of countries with earthquakes	157	168	174	w.wiki/8BRe
4	Количество вулканов по странам (график). Число стран с вулканами Number of volcanoes by country (chart). Number of countries with volcanoes	85	86	86	w.wiki/DxFR
5	Число вулканов России, включая активные и потухшие; генерация карты Number of volcanoes in Russia, including active and extinct ones; map generation	–	322	326	w.wiki/DxEM
6	Список землетрясений с координатами без повторов List of earthquakes with coordinates without duplicates	1958	2055	2172	w.wiki/Dydp
7	Список вулканов на Земле с координатами без повторов List of volcanoes on Earth with coordinates without duplicates	1689	1825	2002	w.wiki/Dyeh

Необходимо пояснение к запросу 6, что значит «без повторов». Дело в том, что в Викиданных одному и тому же землетрясению могут быть приписаны несколько координат (например, координаты эпицентра по разным источникам). Чтобы получить по одной координате на землетрясение, мы используем агрегирующую функцию *SAMPLE*, которая возвращает одно (произвольное) значение из группы возможных (см. код запроса w.wiki/Dydp).

Аналогично «без повторов» получены вулканы с координатами в запросе 7. Также в код добавлено ограничение, что вулканы относятся к Земле, а не к другим астрономическим объектам (см. код запроса w.wiki/Dyeh).

Таким образом, с помощью запросов 6 и 7 из таблицы мы получили из Викиданных список вулканов и землетрясений с географическими координатами. CSV-файлы с этими данными доступны на сайте проекта (https://github.com/componavt/volcano_wikidata/). Эти данные будем использовать в следующем разделе.

ГЕОГРАФИЧЕСКАЯ КЛАСТЕРИЗАЦИЯ НА ОСНОВЕ РАССТОЯНИЯ МЕЖДУ ОБЪЕКТАМИ: ВЫБОР ОПТИМАЛЬНОГО РАДИУСА КЛАСТЕРИЗАЦИИ

Задача состоит в том, чтобы сгруппировать вулканы и землетрясения, находящиеся ближе

некоторой заданной величины `dist_max`. Вулканы и землетрясения считаем объектами одного типа. Максимальное расстояние (радиус кластеризации) – такое, что если вулканы и землетрясения находятся на меньшем расстоянии, то они объединяются в один кластер.

Алгоритм *агломеративной иерархической кластеризации* с перебором радиуса кластеризации (переменная `dist_max`) включает шаги:

1. *Построение KD-дерева* по всем координатам вулканов и землетрясений для быстрого поиска ближайших соседей.
2. *Создание графа*. Для каждого значения `dist_max` из заданного диапазона строится невзвешенный граф. Вулканы и землетрясения – это вершины графа. Ребра добавляются между всеми парами точек, находящимися ближе `dist_max`. Расстояние вычисляется по формуле Хаверсина, учитывающей кривизну Земли.
3. *Выделение компонент связности* построенного графа как групп (кластеров). Вычисление для всех компонент: размера группы и диаметра (расстояние по формуле Хаверсина между максимально удаленными друг от друга вершинами).
4. *Анализ метрик групп*: их количество, среднее и максимальное число объектов в группе, диаметр в километрах (рис. 2).

Алгоритм кластеризации получился быстрым¹ благодаря поиску ближайших землетрясений и вулканов по KD-дереву [7].

Итак, какое нужно выбрать расстояние, чтобы вулканы и землетрясения были связаны? В работе [8] показано, что временная корреляция (от минут до часов) между землетрясениями и извержениями грязевого вулкана наиболее выражена при расстоянии между ними в пределах порядка 100 км при силе землетрясения в 6 и более баллов по шкале Меркалли.

Изменение параметров кластеров в зависимости от максимального допустимого расстояния между вулканами и землетрясениями представлено на рис. 2. На рисунке есть три оси Y: слева – для числа групп и размера наибольшей группы, справа – для среднего размера групп (со шкалой от 3 до 7,5 элемента в группе) и диаметра наибольшей группы (со шкалой от 1 до 4 тыс. км).

Для максимального расстояния выбран крупный шаг в 10 км для всего диапазона 80–160 км, но для участка 120–130 км использован

шаг 1 км, что позволяет выделить на графике локальный изгиб зависимости.

Количество групп на рис. 2 уменьшается достаточно равномерно с увеличением порога кластеризации от 80 до 160 км. На рисунке это кажется не столь равномерным, поскольку мы выбрали неравномерную шкалу по оси X, чтобы растянуть самый интересный участок от 120 до 130 км, где происходят быстрые изменения размера и диаметра самого большого кластера.

На рис. 2 видно, что в диапазоне 127–128 км происходят большие изменения:

- группы вулканов и землетрясений начинают сливаться в один гигантский кластер, размер наибольшей группы увеличивается с 313 до 452 объектов;
- диаметр самой крупной группы скачком возрастает с 2,4 до 4 тыс. км.

При увеличении максимального расстояния от 127 до 128 км появляется суперкластер. Это позволяет определить порог кластеризации (оптимальный радиус) в 127 км – до момента появления очень больших кластеров.

Выясним далее, какие именно кластеры вулканов и землетрясений, существовавшие отдельно при радиусе 127 км, сливаются в общий кластер при 128 км.

ОБЪЕДИНЕНИЕ КЛАСТЕРОВ ВУЛКАНОВ И ЗЕМЛЕТРЯСЕНИЙ В СУПЕРКЛАСТЕР

На рис. 2 видно, что самые большие кластеры при радиусе кластеризации (`dist_max`) равном 127 и 128 км содержат соответственно 313 и 452 элемента. Самая большая группа из 313 элементов не участвует в объединении.

Проанализированы только те группы, которые объединяются при переходе от 127 к 128 км. Найдены две группы, которые на этом шаге формируют суперкластер из 452 элементов:

- группа 1 из 228 объектов (144 вулкана, 84 землетрясения);
- группа 2 из 223 объектов (209 вулканов, 14 землетрясений).

Для визуализации этих групп вулканов и землетрясений на карте (рис. 3) реализованы следующие методы:

1. Алгоритм построения минимального остовного дерева (англ. Minimum Spanning Tree, MST) для выбора и рисования ребер, связывающих элементы кластера.

¹Исходный код программы кластеризации на основе расстояния между объектами доступен онлайн: https://github.com/componavt/volcano_wikidata/blob/main/src/volcano_earthquake_clustering.ipynb.

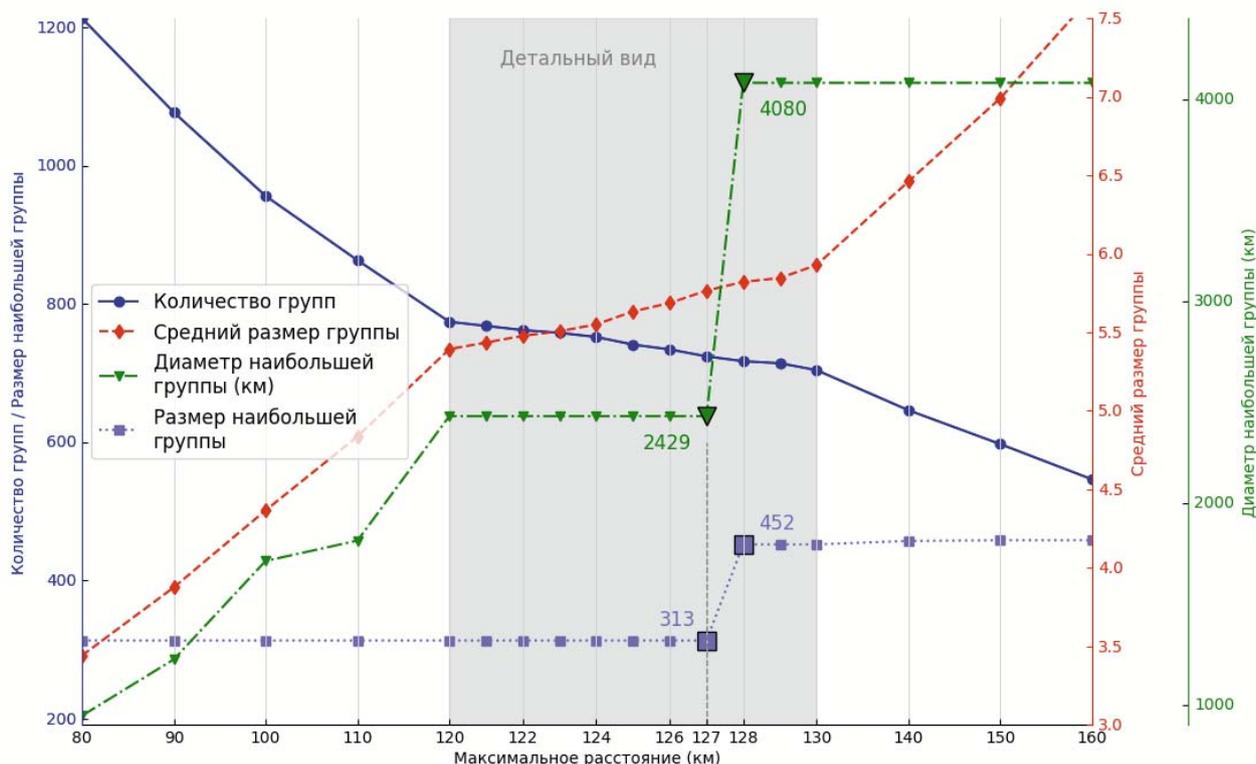


Рис. 2. Зависимость параметров групп от максимального расстояния в кластере между вулканами и землетрясениями

Fig. 2. Cluster metrics as a function of maximum pairing distance (volcano–earthquake)

- Алгоритм жадного покрытия группы кругами с максимальным охватом для минимизации общего числа кругов (эти многочисленные круги радиуса `dist_max`, касаясь друг друга, наглядно показывают, что вершины в них относятся к одному кластеру).
- Метод вычисления центра кластера на основе самой удаленной пары точек. Метод гарантирует, что даже самые удаленные точки окажутся внутри построенного круга.

Интерактивная карта (рис. 3) создается тем же кодом на языке Python¹.

Рисунок 3 полностью построен по Викиданным. То есть известные островные дуги Японская и Курило-Камчатская были выявлены с помощью алгоритма кластеризации на основе расстояния между объектами. Этот результат подтверждает:

- работоспособность предложенного алгоритма,
- достаточную полноту Викиданных для решения таких задач.

Описанный выше алгоритм кластеризации является аналогом алгоритма DBSCAN, который характерен тем, что позволяет выявлять кластеры сложной формы (рис. 3).

Нерешенный вопрос – как оценить качество найденных кластеров? Постараемся приблизиться к ответу в следующем разделе о числе смешанных кластеров, содержащих и вулканы, и землетрясения.

СМЕШАННЫЕ ГРУППЫ В ГЕОГРАФИЧЕСКОЙ КЛАСТЕРИЗАЦИИ НА ОСНОВЕ РАССТОЯНИЯ МЕЖДУ ОБЪЕКТАМИ

Основной целью исследования был поиск взаимосвязи между вулканической и сейсмической активностью, поэтому особый интерес представляют смешанные группы, содержащие оба типа объектов.

На рис. 4 представлена зависимость числа трех возможных групп, получаемых при кластеризации, от максимального допустимого расстояния между вулканами и землетрясениями.

¹Код для генерации интерактивной карты: https://github.com/comonavt/volcano_wikidata/blob/main/src/volcano_earthquake_clustering.ipynb.

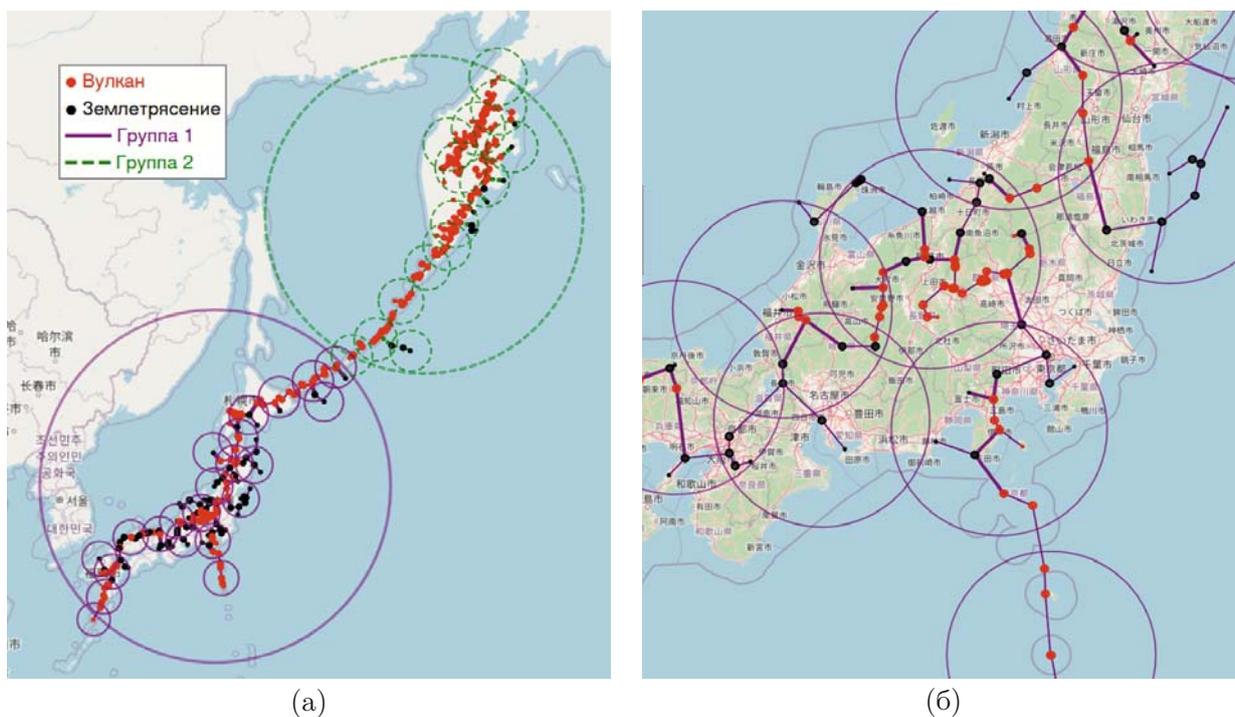


Рис. 3. Два крупных кластера перед объединением – Японская островная дуга (группа 1) и Курило-Камчатская вулканическая дуга (группа 2):

(а) – диаметры групп составляют 2429 км (сплошная линия) и 1680 км (пунктирная линия);

(б) – фрагмент остовного дерева группы 1 (Япония). Толстые линии – ребра MST, соединяющие вулканы и землетрясения. Центральные вершины обозначены крупными точками

Fig. 3. Two large clusters before merging – Japanese island arc (Group 1) and Kuril-Kamchatka volcanic arc (Group 2):

(a) – cluster diameters are 2429 km (solid line) and 1680 km (dashed line);

(б) – a fragment of the spanning tree for Group 1 (Japan). MST edges connecting volcanoes and earthquakes are shown with thicker lines. Central vertices are slightly enlarged

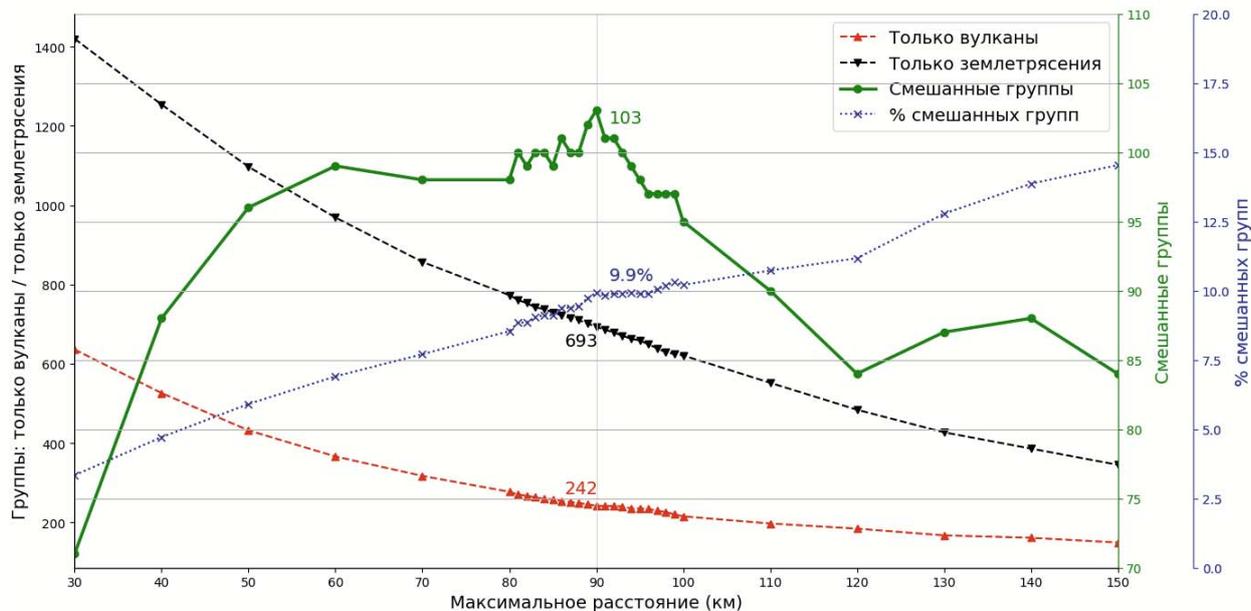


Рис. 4. Зависимость числа вулканических, сейсмических и смешанных групп от порога кластеризации

Fig. 4. Relationship between cluster counts (volcanic/seismic/mixed) and the distance threshold parameter

Максимум смешанных кластеров (103 группы, 9,9%) наблюдается при `dist_max` = 90 км, что может отражать типичный радиус взаимодействия (*spatial correlation distance*) (~100 км), зафиксированный для грязевых вулканов [8], где землетрясения статистически значимо провоцируют извержения.

При различных значениях максимального расстояния преобладают группы, содержащие только землетрясения. Например, при `dist_max` = 90 км кластеризация дает 693 сейсмические группы (66,8%). Это указывает на более локализованное распределение землетрясений.

ЗАКЛЮЧЕНИЕ

Разработаны SPARQL-скрипты для получения информации о вулканах и землетрясениях из Викиданных. Разработана программа кластеризации и визуализации на карте ближайших вулканов и землетрясений. Разработанные алгоритмы и собранные данные доступны онлайн (https://github.com/componavt/volcano_wikidata).

Предложенный алгоритм географической кластеризации на основе расстояния между объектами аналогичен DBSCAN, но адаптирован для пространственных данных:

- использует единое пороговое расстояние (`dist_max`) вместо плотностных параметров DBSCAN (радиус окрестности и минимальное число точек для ядра);
- как и DBSCAN, он способен выявлять кластеры произвольной формы. Выявлены островные дуги Японская и Курило-Камчатская (рис. 3);
- не требует равномерного распределения объектов, критичного для классического DBSCAN.

Проведены анализ и сравнение радиусов кластеризации для разработанного алгоритма кластеризации. Получены два ключевых значения:

- 90 км – оптимальное расстояние, при котором доля смешанных кластеров (вулканы + землетрясения) максимальна (103 группы). Этот результат согласуется с известными данными о радиусе взаимодействия (~100 км) для грязевых вулканов [8], но впервые расширен на обычные вулканы.
- 127–128 км – диапазон, где происходит качественное изменение структуры,

а именно: слияние Японской и Курило-Камчатской дуг в единый суперкластер с резким увеличением диаметра (2,4 → 4 тыс. км), что свидетельствует о потере географической интерпретируемости.

Отмеченный выше максимум при пороге в 90 км – это 103 смешанные группы, которые составляют всего 10% от числа всех групп. В будущей работе стоит модифицировать алгоритм кластеризации, чтобы в нем была больше доля групп смешанного типа.

У нашего алгоритма кластеризации есть недостаток – это сильная чувствительность к выбору `dist_max`. В диапазоне 120–130 км (рис. 2) даже 1 км изменения порога приводит к качественному скачку: слиянию изолированных кластеров в суперкластеры.

В работе мы исследовали ту часть Викиданных, которая описывает вулканы и землетрясения. Рассмотрены открытые базы данных вулканов GVP (*Global Volcanism Program*) и LAMEVE (*Large Magnitude Explosive Volcanic Eruptions*), поскольку Викиданные частично основываются на этих базах и объединяют их с помощью внешних идентификаторов.

Разработанные SPARQL-скрипты автоматизируют извлечение и предварительную очистку данных, а именно, обеспечивают:

- фильтрацию дубликатов координат (использование агрегирующей функции `SAMPLE` сократило число повторяющихся точек на 5,2%);
- стандартизацию формата данных для последующей кластеризации;
- возможность регулярного обновления выборки при изменении Викиданных.

Дубликаты координат связаны с такой особенностью Викиданных, как многозначность. На примере объекта *Antsiferov Island* (Q614531) показано наличие двух альтернативных значений координат (расхождение ~ 390 м). Подробнее о наличии множества значений, связанных с техническими ошибками или различными идеологиями и точками зрения, см. в работе [10, с. 2–5].

Викиданные служат не просто справочником, а полноценной платформой для исследований. Они позволяют получать пространственные данные для объектов заданного типа за несколько запросов, обеспечивают прозрачность источников через систему внешних идентификаторов.

Викиданные позволяют анализировать пространственные связи вулканов и землетрясений (реализовано в Python-программе), но временной анализ ограничен неполнотой датировок: только 3,3% вулканов имеют данные об извержениях. Именно объект «извержение», а не «вулкан» может иметь в Викиданных дату. За 2023–2025 годы число датированных извержений в Викиданных выросло с 7 до 68, что свидетельствует об активном пополнении этих данных.

Работа демонстрирует, что Викиданные, несмотря на ограничения, могут служить достоверным источником для исследования пространственных взаимосвязей между вулканами и землетрясениями. Выявленные кластеры соответствуют известным геологическим структурам (Японская и Курило-Камчатская дуги), что подтверждает адекватность данных. Эти результаты демонстрируют, что краудсорсинговые базы данных могут дополнять традиционные научные ресурсы при исследовании глобальных геофизических закономерностей.

ЛИТЕРАТУРА

1. Крижановский А. А., Балакирева М. С., Меньшикова Е. А., Паренченков Е. О., Потес А. С., Трубина Е. Д., Обрегон А. Программирование Викиданных. Петрозаводск: ПетрГУ, 2024. 186 с. URL: <https://commons.wikimedia.org/?curid=146556492> (дата обращения: 04.05.2025).

2. Bebbington M. S., Marzocchi W. Stochastic models for earthquake triggering of volcanic eruptions // *J. Geophys. Res.: Solid Earth*. 2011. Vol. 116. Art. B05204. doi: 10.1029/2010JB008114

3. Caricchi L., Townsend M., Rivalta E., Namiki A. The build-up and triggers of volcanic eruptions // *Nat. Rev. Earth Environ*. 2021. No. 2. P. 458–476. doi: 10.1038/s43017-021-00174-8

4. Costa F., Widiwijayanti C., Zar Win Nang T., Fajiculay E., Espinosa-Ortega T., Newhall C. WOVOdat – the global volcano unrest database aimed at improving eruption forecasts // *Disast. Prevent. Manag.* 2019. Vol. 28, no. 6. P. 738–751. doi: 10.1108/DPM-09-2019-0301

5. Crosweller H. S., Arora B., Brown S. K., Cottrell E., Deligne N. I., Ortiz Guerrero N., Hobbs L., Kiyosugi K., Loughlin S. C., Lowndes J., Nayembil M., Siebert L., Sparks R. S. J., Takarada S., Venzke E. Global database on large magnitude explosive volcanic eruptions (LaMEVE) // *J. Appl. Volcanol.* 2012. Vol. 1. Art. 4. doi: 10.1186/2191-5040-1-4

6. Eggert S., Walter T. R. Volcanic activity before and after large tectonic earthquakes: Observations

and statistical significance // *Tectonophysics*. 2009. No. 471. P. 14–26. doi: 10.1016/j.tecto.2008.10.003

7. Maneewongvatana S., Mount D. M. Analysis of approximate nearest neighbor searching with clustered point sets. 1999. arXiv preprint cs/9901013. URL: <https://arxiv.org/pdf/cs/9901013> (дата обращения: 04.05.2025).

8. Mellors R., Kilb D., Aliyev A., Gasanov A., Yetirmishli G. Correlations between earthquakes and large mud volcano eruptions // *J. Geophys. Res.* 2007. Vol. 112. Art. B04304. doi: 10.1029/2006JB004489

9. Papale P. Global time-size distribution of volcanic eruptions on Earth // *Scientific Reports*. 2018. Vol. 8. Art. 6838. doi: 10.1038/s41598-018-25286-y

10. Santos V., Schwabe D., Lifschitz S. Can you trust Wikidata? // *Semantic Web*. 2024. No. 1. P. 1–22. doi: 10.3233/SW-243577

11. Seropian G., Kennedy B. M., Walter T. R., Ichihara M., Jolly A. D. A review framework of how earthquakes trigger volcanic eruptions // *Nat. Commun.* 2021. Vol. 12. Art. 1004. doi: 10.1038/s41467-021-21166-8

12. Siebert L., Simkin T., Kimberly P. *Volcanoes of the World*. 3rd ed. Berkeley: University of California Press, 2010. 568 p.

13. Simkin T. Terrestrial volcanism in space and time // *Annu. Rev. Earth Planet. Sci.* 1993. Vol. 21, no. 1. P. 427–452. doi: 10.1146/annurev.ea.21.050193.002235

14. Walter T. R., Amelung F. Volcano-earthquake interaction at Mauna Loa volcano, Hawaii // *J. Geophys. Res.: Solid Earth*. 2006. Vol. 111. Art. B05204. doi: 10.1029/2005JB003861

15. Zhao F. A Systematic Review of Wikidata in Digital Humanities Projects // *Dig. Scholarsh. Humanit.* 2023. Vol. 38, iss. 2. P. 852–874. doi: 10.1093/llc/fqac083

REFERENCES

1. Krizhanovsky A. A., Balakireva M. C., Menshikova E. A., Parenchenkov E. O., Potes A. S., Trubina E. D., Obregon A. Programming Wikidata. Petrozavodsk: PetrSU; 2024. 186 p. (In Russ.). URL: <https://commons.wikimedia.org/?curid=146556492> (accessed: 04.05.2025).

2. Bebbington M. S., Marzocchi W. Stochastic models for earthquake triggering of volcanic eruptions. *J. Geophys. Res.: Solid Earth*. 2011;116:B05204. doi: 10.1029/2010JB008114

3. Caricchi L., Townsend M., Rivalta E., Namiki A. The build-up and triggers of volcanic eruptions. *Nat. Rev. Earth Environ*. 2021;2:458–476. doi: 10.1038/s43017-021-00174-8

4. Costa F., Widiwijayanti C., Zar Win Nang T., Fajiculay E., Espinosa-Ortega T., Newhall C.

WOVOdat – the global volcano unrest database aimed at improving eruption forecasts. *Disast. Prevent. Manag.* 2019;28(6):738–751. doi: 10.1108/DPM-09-2019-0301

5. *Crosweller H. S., Arora B., Brown S. K., Cottrell E., Deligne N. I., Ortiz Guerrero N., Hobbs L., Kiyosugi K., Loughlin S. C., Lowndes J., Nayembil M., Siebert L., Sparks R. S. J., Takarada S., Venzke E.* Global database on large magnitude explosive volcanic eruptions (LaMEVE). *J. Appl. Volcanol.* 2012;1:4. doi: 10.1186/2191-5040-1-4

6. *Eggert S., Walter T. R.* Volcanic activity before and after large tectonic earthquakes: Observations and statistical significance. *Tectonophysics.* 2009;471:14–26. doi: 10.1016/j.tecto.2008.10.003

7. *Maneewongvatana S., Mount D. M.* Analysis of approximate nearest neighbor searching with clustered point sets. arXiv preprint cs/9901013. 1999. URL: <https://arxiv.org/pdf/cs/9901013> (accessed: 04.05.2025).

8. *Mellors R., Kilb D., Aliyev A., Gasanov A., Yetirmishli G.* Correlations between earthquakes and large mud volcano eruptions. *Geophys. Res.* 2007;112:B04304. doi: 10.1029/2006JB004489

9. *Papale P.* Global time-size distribution of volcanic eruptions on Earth. *Scientific Reports.* 2018;8:6838. doi: 10.1038/s41598-018-25286-y

10. *Santos V., Schwabe D., Lifschitz S.* Can you trust Wikidata? *Semantic Web.* 2024;1:1–22. doi: 10.3233/SW-243577

11. *Seropian G., Kennedy B. M., Walter T. R., Ichihara M., Jolly A. D.* A review framework of how earthquakes trigger volcanic eruptions. *Nat. Commun.* 2021;12:1004. doi: 10.1038/s41467-021-21166-8

12. *Siebert L., Simkin T., Kimberly P.* Volcanoes of the World. 3rd ed. Berkeley: University of California Press; 2010. 568 p.

13. *Simkin T.* Terrestrial volcanism in space and time. *Annu. Rev. Earth Planet. Sci.* 1993;21(1):427–452. doi: 10.1146/annurev.earth.21.050193.002235

14. *Walter T. R., Amelung F.* Volcano-earthquake interaction at Mauna Loa volcano, Hawaii. *J. Geophys. Res.: Solid Earth.* 2006;111:B05204. doi: 10.1029/2005JB003861

15. *Zhao F.* A Systematic Review of Wikidata in Digital Humanities Projects. *Dig. Scholarsh. Humanit.* 2023;38(2):852–874. doi: 10.1093/llc/fqac083

Поступила в редакцию / received: 05.05.2025; принята к публикации / accepted: 23.05.2025.
Автор заявляет об отсутствии конфликта интересов / The author declares no conflict of interest.

СВЕДЕНИЯ ОБ АВТОРЕ:

Крижановский Андрей Анатольевич
канд. техн. наук, ведущий научный сотрудник
e-mail: andew.krizhanovsky@gmail.com

CONTRIBUTOR:

Krizhanovsky, Andrew
Cand. Sci. (Tech.), Leading Researcher