

АНАЛИЗ УЧАСТИЯ ПРЕПОДАВАТЕЛЕЙ ВУЗА В НАУЧНЫХ МЕРОПРИЯТИЯХ С ПОМОЩЬЮ ELASTICSEARCH

А. В. ПРУЦКОВ

Аннотация. Анализ данных является основой для принятия управленческих решений. Разработаны различные программы, позволяющие автоматизировать анализ данных. Одним из видов анализа данных является агрегация. Наиболее разнообразные способы агрегации данных по сравнению с другими программами предоставляет программа Elasticsearch (ES) из пакета программ Elastic Stack. ES – это информационно-поисковая система, одной из функций которой является агрегация данных. Цель работы – продемонстрировать возможности ES в агрегации данных о планах участия преподавателей вуза в научных мероприятиях. С помощью агрегаций термов, количества уникальных значений, количества значений и гистограммы выявлены: 1) подразделения вуза с наибольшим количеством преподавателей, подавших заявки на участие в мероприятиях; 2) мероприятия с наибольшим количеством заявок в одном из подразделений; 3) мероприятия с наибольшим количеством заявок, проводимые в вузе и вне вуза; 4) распределение количества преподавателей в зависимости от числа запланированных ими участия в мероприятиях. По результатам агрегаций автор сделал выводы о неравномерности распределения заявок по подразделениям и предпочтении участия в мероприятиях, организованных вузом, и другие. Запросы агрегаций имеют пояснения. Проведенная агрегация данных показывает недооценку ES при решении этой задачи.

Ключевые слова: Elasticsearch; Elastic Stack; анализ данных; агрегация данных; запросы агрегации; преподаватели; университет; конференция.

Цитирование: Пруцков А. В. Анализ участия преподавателей вуза в научных мероприятиях с помощью Elasticsearch // СИИТ. 2025. Т. 7, № 4(23). С. 38–48. EDN [JKRJQA](#).

ВВЕДЕНИЕ

Данные анализируются не ради анализа, а ради выявления скрытых закономерностей в этих данных. Скрытые закономерности позволяют системно исследовать предметную область, из которой эти данные были получены. При анализе элементы данных могут обрабатываться следующими способами:

- преобразование (перевод градусов в радианы);
- изменение порядка (сортировка);
- отбор по условию, в том числе с удалением из набора;
- подсчет однозначных метрик (среднее арифметическое, минимум, максимум и др.);
- группировка.

Перечисленные способы могут применяться в различных допустимых комбинациях для агрегации данных. Агрегация данных – это часть их анализа, состоящая в получении производных данных.

Агрегируем следующий набор данных.

НАБОР ИСХОДНЫХ ДАННЫХ

Согласно распоряжению ректора Рязанского государственного радиотехнического университета имени В. Ф. Уткина (РГРТУ, далее – «вуз») преподавателям было необходимо внести в электронную таблицу заявки на участие в научных мероприятиях на 2024/2025 учебный год. Таблица находится в открытом доступе на известном веб-сервисе. Эти заявки были частью плана научной работы вуза.

Данные в таблице имеют следующие поля; в скобках указаны имена некоторых полей:

- номер записи (начинается с 1);
- наименование мероприятия (event_name);
- тип мероприятия (конференция; симпозиум; семинар; форум; другое);

- уровень мероприятия (международный; региональный; всероссийский; другое);
- форма участия (очная; заочная; очно-заочная; дистанционная; презентация доклада);
- ориентировочные сроки проведения мероприятия;
- фактические сроки проведения мероприятия;
- подразделения (кафедры и НИИ вуза, department);
- сотрудники, участвующие в мероприятии (перечисляются один или несколько сотрудников подразделения, staff).

Перед агрегацией данные были предобработаны следующим образом:

- унифицированы названия мероприятий;
- устранено дублирование заявок;
- добавлено поле признака вуза-организатора (РГРТУ / Не РГРТУ, university).

После предобработки набор данных стал иметь следующие статистические показатели:

- записей данных: 161;
- мероприятий: 42;
- подразделений: 20;
- преподавателей: 161;
- заявок на участие: 328.

Под заявкой на участие понимается плановое участие одного преподавателя в одном мероприятии.

Участие преподавателей вузов в конференциях анализировалось в исследованиях по различным направлениям:

- Научно-деятельностное направление. Участие преподавателей в региональных и международных конференциях, семинарах является составляющей научной деятельности в регионе [Пуз14]. На веб-сайтах омских вузов размещается информация об участии преподавателей в научных мероприятиях, проводимых вузами Омской области и вузами других регионов России. В [Мир19] проведен эконометрический анализ научно-исследовательской деятельности (НИД) двух региональных вузов. Участие в конференциях является одним из факторов. Сделан вывод о том, что НИД вузов имеет стихийный характер, а не сознательно-управляемый.

- Психологическое направление. Профессора на международных междисциплинарных научных конференциях меньше контактируют, чем другие участники конференции [Gén19]. Причиной этого является отсутствие необходимости у профессоров заводить новые контакты. В [Lut24] выявлено, что ученые-девушки контактируют с профессорами обоих полов активней, чем ученые-юноши. По результатам собеседований с участниками конференции было рекомендовано организаторам конференций учитывать не только желания участников, но и их нужды [Row18].

Отношение преподавателей к участию в научных конференциях бывает пренебрежительным, утилитарным, имеет злоупотребления в виде предоставления материалов не по профилю конференции, «увлечения» соавторством [Бал05]. Автор статьи [Бал05] заключает, что «год от года на научных конференциях все меньше и меньше появляется ярких, интересных докладов с оригинальными идеями и действительно новыми практическими подходами».

Проанализируем данные об участии преподавателей вуза в научных мероприятиях.

ВЫБОР ПРОГРАММЫ АГРЕГАЦИИ ДАННЫХ

Агрегацию данных можно автоматизировать. Степень автоматизации агрегации определяется функциями инструментария. Одним из инструментариев являются программы обработки и анализа данных. Кроме табличного процессора Microsoft Excel, популярными программами являются следующие (на основе [Sza23]). В описаниях указаны ссылки на публикации с порядком обработки и анализа данных в этих программах.

R – это язык статистической обработки данных [Fre19, Kab11]. Язык имеет различные среды разработки. Возможности языка постоянно расширяются новыми библиотеками,

поэтому язык R может обрабатывать текст, строить графики и диаграммы. Мы подробно исследовали возможности этого языка в [Ава24].

Python – это полноценный язык программирования с расширяемыми возможностями за счет многочисленных библиотек (например, scikit-learn) [Bri17, Gra15]. Является наиболее популярным языком программирования для анализа данных.

SQL – язык изменения и получения данных и структур таблиц из реляционных баз данных [Fos21, Sha22].

Apache Hadoop – программная инфраструктура для обработки «больших данных» [Kha22], включающая распределенную файловую систему HDFS и среду распределенной обработки данных MapReduce.

Apache Spark – программа управления «большими данными» и их параллельной обработки [Per20].

Microsoft Power BI [Kni22] и Tableau [Mil22] – программы для анализа бизнес-данных: исследования наборов данных из различных источников и их интерактивного графического представления.

Qlik Sense [Vit18], RapidMiner [Ram23] и KNIME [Gre20] – программы для интеллектуального анализа данных. RapidMiner имеет визуальную среду для разработки порядка обработки данных [For23].

Анализ данных в Microsoft Excel, на языках R и Python сравнивается в [Lee22a, Lee22b, Mou21].

Анализ бизнес-данных (business intelligence, BI) заключается в построении графического представления данных и их статистической обработки, на основе которых компании принимают управленческие решения [Sah21].

Программы для анализа бизнес-данных предназначены для их графического представления и совместного использования, в то время как языки R/Python предлагают различные функции анализа и возможность учета требований конкретной задачи. В некоторых случаях можно использовать программы обоих типов, распределяя их возможности на различные этапы анализа [Sco23].

Перечисленные программы можно использовать для анализа исходного набора данных. Однако многие виды агрегации, имеющиеся в программе Elasticsearch (ES), в них отсутствуют. ES нечасто используется для агрегации данных, например для агрегации уникальных пользователей и связанных с ними записей протоколов работы порталов пространственных данных [Li19], показателей модуля автомасштабирования [Cho18].

ЦЕЛЬ РАБОТЫ

Целью работы является демонстрация возможности агрегации данных об участии преподавателей в научных конференциях программой ES.

ПРОГРАММА ELASTICSEARCH И ПОРЯДОК АНАЛИЗА ДАННЫХ В НЕЙ

ES – это документо-ориентированная информационно-поисковая система (ИПС). Элементы данных хранятся в документах. Документы хранятся в индексе ES.

Программа ES входит в пакет программ Elastic Stack вместе с другими программами:

- Kibana обеспечивает графический интерфейс для отправки запросов и приемом ответов в формате JSON, управление индексами;
- Logstash позволяет загружать данные в ES из различных источников.
- Преимуществами ES являются следующие:
- широкие возможности агрегации данных;
- бесплатное использование (при некоторых ограничениях);
- простота установки (скачать–распаковать–запустить).

Порядок анализа данных состоит из следующих шагов:

- загрузить данные в индекс ES с помощью программы Logstash; для это необходимо написать конфигурацию загрузки;
- выполнить запросы и получить ответы в формате JSON с помощью программы Kibana;
- преобразовать ответы в формате JSON в табличную форму.

Последние версии программы Kibana позволяют загружать данные из файлов в формате CSV в индекс ES, однако в этом исследовании это сделать не получилось.

Для преобразования JSON в табличную форму использовались специализированные веб-сервисы. Веб-сервисы можно найти с помощью запроса «convert json to csv» любой ИПС.

АГРЕГАЦИЯ В ПРОГРАММЕ ELASTICSEARCH

Агрегации в программе ES бывают трех видов: 1) по группам; 2) статистические; 3) конвейерные. Пусть индекс программы Elasticsearch содержит документы:

$$D = \{d_1, d_2, \dots, d_{ND}\}.$$

Агрегация по группам можно записать формулой:

$$A = \{a_1, a_2, \dots, a_{NA}\}.$$

Каждая группа включает условие отбора документов и подмножество документов:

$$a_i = \langle P_i(d), G_i \rangle, i = 1, 2, \dots, NA;$$

$$(d_j \in D) \wedge (d_j \in G_i) \wedge P(d_j), j = 1, 2, \dots, ND,$$

где $P_i(d)$ – условие отбора документов, предикат; $G_i \subseteq D$. При этом

$$G_i \cap G_j = \emptyset, j = 1, 2, \dots, NA, i \neq j;$$

$$\forall d((d \in D) \wedge \overline{P_1(d)} \wedge \overline{P_2(d)} \wedge \dots \wedge P_k(d) \wedge \dots \wedge \overline{P_{NA}(d)}), k = 1, 2, \dots, NA.$$

Агрегации по группам могут быть вложенными: агрегация применяется не только к документам множества D , но и к документам множеств G_i .

Значениями статистических агрегаций является число t (сумма, минимум, максимум и т. п.), вычисленное по значениям полей документов: $t = f(d_1, d_2, \dots, d_{ND})$.

Конвейерные агрегации выполняют постобработку: вычисление значений по группам, сортировку и отбор групп.

РЕЗУЛЬТАТЫ АГРЕГАЦИИ НАБОРА ДАННЫХ И ВЫВОДЫ

С помощью ES были получены следующие результаты агрегации данных набора. Семь подразделений с наибольшим количеством заявок на участие (табл. 1) подали 60% всех заявок.

Таблица 1

Количество преподавателей, подавших заявки на участие в научных мероприятиях, и количество заявок по подразделениям

Кафедра	Количество преподавателей, подавших заявки	Количество заявок на участие
ВПИМ	18	41
КТ	13	35
МНЭЛ	8	31
ИИБМТ	25	25
ОиЭФ	11	23
РТУ	6	22
АИТП	4	21

Исследуем заявки на участие преподавателей кафедры ВПМ. Тремя мероприятиями с наибольшим количеством заявок от преподавателей этой кафедры являются следующие (табл. 2).

Таблица 2

**Три мероприятия с наибольшим количеством заявок
от преподавателей кафедры ВПМ**

Мероприятие	Количество заявок на участие
Новые информационные технологии в научных исследованиях (НИТ)	13
Современные технологии в науке и образовании (СТНО-2025)	9
Новые информационные технологии в образовании 2025. Интеграция ИТ-индустрии и системы образования на базе технологических решений 1С – стратегии роста	3

Выясним по два мероприятия с наибольшим количеством заявок, проводимых в вузе и вне вуза (табл. 3).

Таблица 3

**Мероприятия с наибольшим количеством заявок,
проводимых в вузе и вне вуза**

Количество заявок на участие от одного преподавателей	Вуз-организатор	Количество преподавателей
СТНО-2025	РГРТУ	101
НИТ	РГРТУ	37
Цифровая обработка сигналов (DSPА-2025)	Не РГРТУ	10
Новые информационные технологии в образовании 2025. Интеграция ИТ-индустрии и системы образования на базе технологических решений 1С – стратегии роста	Не РГРТУ	3

Рассмотрим распределение преподавателей по количеству поданных ими заявок (табл. 4). Данные для последнего столбца были получены с помощью Microsoft Excel.

Таблица 4

Распределение преподавателей по количеству поданных ими заявок

Количество заявок на участие от одного преподавателей	Количество преподавателей	Доля
7	1	0,6%
6	3	1,9%
5	11	6,8%
4	17	10,6%
3	15	9,3%
2	21	13,0%
1	93	57,8%

Исходный набор данных может быть агрегирован и далее. Ограничим результаты приведенными выше.

Из полученных результатов можно сделать следующие выводы:

- преподаватели планируют участвовать в разнообразных научных мероприятиях;
- преподаватели планируют участвовать чаще в научных мероприятиях, проводимых вузом, но не вне вуза;
- заявки на участие распределяются среди подразделений очень неравномерно;
- большинство преподавателей планируют участвовать в одном научном мероприятии.

Причиной таких выводов являются недостаточная научная деятельность преподавателей, что в свою очередь вызвано отсутствием научной работы, отсутствием контроля за научной деятельностью преподавателей со стороны руководства вуза. Научная деятельность преподавателей может быть повышена за счет введения норм участия в конференциях вне вуза (а также увеличения числа научных публикаций) при подаче документов на конкурс, помощи руководства при поиске научно-исследовательских работ.

ЗАПРОСЫ АГРЕГАЦИИ ДАННЫХ

Представленные результаты были получены в ответ на запросы ES. Запросы записаны в формате JSON.

Статистические показатели предобработанного набора данных были получены запросами, аналогичным запросу (листинг 1). Запрос агрегации количества преподавателей, подавших заявки на участие в научных мероприятиях, является запросом агрегации количества уникальных значений (cardinality) по полю staff (строки 5–6).

Листинг 1

Запрос агрегации количества преподавателей, подавших заявки на участие в научных мероприятиях

```
001 GET /conferences/_search?filter_path=aggregations
002 {
003   "aggs": {
004     "staff_count": {
005       "cardinality": {
006         "field": "staff.keyword"
007       } } } }
```

Количество преподавателей и их заявок по подразделениям (табл. 1) было получено запросом (листинг 2). В запросе используется агрегация термов (terms) и вложенные агрегации. Агрегация термов (строки 5–10) группирует документы в зависимости значения термина. Данные агрегации сортируются по убыванию значения вложенной агрегации (строки 8–10) и ограничены 7 термами (строка 7). Вложенными агрегациями являются агрегация количества значений в поле (value_count, строки 14–16) и агрегация количества уникальных значений (строки 19–20). Первая вложенная агрегация подсчитывает количество заявок, а вторая – количество преподавателей по каждому подразделению.

Листинг 2

Запрос агрегации по подразделениям количества преподавателей, подавших заявки на участие в научных мероприятиях, и количества заявок

```
001 GET /conferences/_search?filter_path=aggregations
002 {
003   "aggs": {
004     "by_department": {
005       "terms": {
006         "field": "department.keyword",
```

```

007         "size": 7,
008         "order": {
009             "event_request_count": "desc"
010         },
011     },
012     "aggs": {
013         "event_request_count": {
014             "value_count": {
015                 "field": "staff.keyword"
016             }
017         },
018         "staff_count": {
019             "cardinality": {
020                 "field": "staff.keyword"
021             }
022         }
023     }
024 }

```

Данные о мероприятиях, в которых планируют участвовать преподаватели кафедры ВПМ, (табл. 2) были получены запросом (листинг 3). В запросе используются агрегация термов (строки 12–17) и вложенная в нее агрегация количества уникальных значений (строки 21–22). В запросе используется запрос поиска термов (term, строки 4–8) для отбора документов по кафедре ВПМ.

Мероприятия с наибольшим количеством заявок, проводимых в вузе и вне вуза (табл. 3) были получены запросом (листинг 4). Запрос состоит из трех агрегаций: по вузу-организатору (строки 5–7), по мероприятию (строки 10–13) и по преподавателям (строки 16–17).

Распределение преподавателей по количеству поданных ими заявок (табл. 4) не может быть получено одним запросом. Была использована функция трансформации данных (листинг 5) – агрегации данных (строки 18–20) и запись их в другой индекс. Собственно данные агрегируются запросом построения гистограммы (листинг 6).

Приведенные запросы просты в понимании и написании (после некоторой практики). Для углубления знаний об агрегации данных в ES и работе этой системы существуют следующие справочные материалы:

- официальная документация к ES с примерами на английском языке, находящаяся по адресу elastic.co/guide/index.html; документация подробна и содержит поясняющие примеры;
- книги; наиболее подходящие для начального освоения комплекса программ Elastic Stack книги выявлены в [Пру23];
- форумы пользователей ES; наиболее представительные форумы находятся по адресам stackoverflow.com и discuss.elastic.co в сети Интернет.

Листинг 3

Запрос агрегации по мероприятиям количества заявок на участие от преподавателей кафедры ВПМ

```

001 GET /conferences/_search?filter_path=aggregations
002 {
003     "query": {
004         "term": {
005             "department.keyword": {
006                 "value": "ВПМ"
007             }
008         }
009     },
010     "aggs": {

```

```

011     "by_event": {
012         "terms": {
013             "field": "event_name.keyword",
014             "size": 3,
015             "order": {
016                 "event_request_count": "desc"
017             }
018         },
019         "aggs": {
020             "event_request_count": {
021                 "cardinality": {
022                     "field": "staff.keyword"
023                 }
024             }
025         }
026     }
027 }

```

Листинг 4

**Запрос агрегации мероприятия с наибольшим количеством заявок,
проводимых в вузе и вне вуза**

```

001 GET /conferences/_search?filter_path=aggregations
002 {
003     "aggs": {
004         "by_university": {
005             "terms": {
006                 "field": "university.keyword",
007             },
008             "aggs": {
009                 "by_event": {
010                     "terms": {
011                         "field": "event_name.keyword",
012                         "size": 2
013                     },
014                     "aggs": {
015                         "staff_count": {
016                             "cardinality": {
017                                 "field": "staff.keyword"
018                             }
019                         }
020                     }
021                 }
022             }
023         }
024     }
025 }

```

Листинг 5

**Запрос трансформации данных в распределение преподавателей
по количеству поданных ими заявок**

```

001 PUT _transform/event_count_by_staff
002 {
003     "description": "Event count by staff",
004     "source": {
005         "index": "conferences"
006     },
007     "dest": {
008         "index": "conferences_staff"
009     },
010     "pivot": {
011         "group_by": {
012             "staff": {

```



```

013         "terms": {
014             "field": "staff.keyword"
015         }
016     },
017     "aggregations": {
018         "event_count": {
019             "cardinality": {
020                 "field": "event_name.keyword"
021             }
022         }
023     }
024 }

```

Листинг 6

Запрос агрегации распределения преподавателей по количеству поданных ими заявок

```

001 GET /conferences_staff/_search?filter_path=aggregations
002 {
003     "aggs": {
004         "staff_event_count_histogram": {
005             "histogram": {
006                 "field": "event_count",
007                 "interval": 1
008             }
009         }
010     }
011 }

```

ЗАКЛЮЧЕНИЕ

В статье продемонстрированы примеры агрегации данных с помощью ES. Разнообразие агрегаций ES позволяет получить выводы, которые будут использоваться при принятии управленческих решений. ES чаще используется как ИПС или хранилище протоколов работы программ [Pru25]. Ее возможности агрегации данных недооценены. Освоить ES стоит еще и потому, что это динамично развивающаяся ИПС, которая теперь имеет функции интеллектуального анализа данных и векторного поиска.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- [Пуз14] Пузина Н. В. Анализ научной активности преподавателей омских вузов // Вестник Сибирского института бизнеса и информационных технологий. 2014. № 2 (10). С. 99-104. EDN: [SNRXGH](#). [[Puzina N.V. "Analysis of scientific activity of Omsk universities staff" // Herald of Siberian Institute of Business and Information Technologies. 2014, no. 2 (10), pp. 99–104. EDN: [SNRXGH](#). (In Russian).]]
- [Мир19] Миролюбова А. А., Ксенофонтова О. Л. Научно-исследовательская деятельность вузов: сравнительный анализ и оценка результатов // Известия вузов. Серия «Экономика, финансы и управление производством». 2019. № 04 (42). С. 131-138. EDN: [QUAPSI](#). [[Miroljubova A. A., Ksenofontova O. L. "Scientific activity of universities: comparing, analysis, and estimation of results" // Proceedings of Universities. The Economics, Finance, and Production Management Serie. 2019. No. 04 (42), pp. 131–138. EDN: [QUAPSI](#). (In Russian).]]
- [Gén19] Géniois M. et al., "Building Connections: How Scientists Meet Each Other During a Conference" // arXiv Preprint, 2019. p. 1901.01182. DOI: [10.48550/arXiv.1901.01182](#).
- [Lut24] Lutter M. et al., "Who Talks to the Prof? Gender Differences in Interaction with Senior Scholars at Four Academic Conferences" // Preprint. 2024. DOI: [10.31235/osf.io/9pzyrn](#).
- [Row18] Rowe N. "When You Get What You Want, But Not What You Need": The Motivations, Affordances and Shortcomings of Attending Academic/Scientific Conferences" // International Journal of Research in Education and Science. 2018. Vol. 4, No. 2, pp. 714-729. DOI: [10.21890/ijres.438394](#).
- [Бал05] Балабан В.А. Научно-исследовательская работа и участие преподавателей вуза в научных конференциях // Известия Дальневосточного федерального университета. Экономика и управление. 2005. № 4. С. 125-130. EDN: [JXUHRX](#). [[Balaban V. A. "Research Work and university staff participant in scientific conferences" // Proceedings of the Far-Eastern Federal University. Economics and Management. 2005. No. 4, pp. 125–130. EDN: [JXUHRX](#). (In Russian).]]
- [Sza23] Szasz B. Data Analytics Essentials You Always Wanted to Know. Vibrant Publishers, 2023.
- [Fre19] Freeman M., Ross J. Programming Skills for Data Science. Start Writing Code to Wrangle, Analyze, and Visualize Data with R. Addison-Wesley, 2019.
- [Kab11] Kabacoff R. R in Action. Data Analysis and Graphics with R. Manning, 2011.

- [Ава24] Авачева Т. Г., Пруцков А. В., Медведева О. В. Современные технологии решения задач медицинской статистики // Современные проблемы здравоохранения и медицинской статистики. 2024. № 2. С 292-303. DOI: [10.24412/2312-2935-2024-2-292-303](#). EDN: [ZGNPWH](#). [[Avacheva T. G., Prutzkow A. V., Medvedeva O. V. "Modern technologies to solve problems in medical statistics" // Actual Problems of Public Health Medical Statistics. 2024. No. 2, pp. 292–303. DOI: [10.24412/2312-2935-2024-2-292-303](#). (In Russian).]]
- [Bri17] Brink H. et al. Real-World Machine Learning. Manning, 2017.
- [Gra15] Gras J. Data Science from Scratch. O'Reilly, 2015.
- [Fos21] Foster I. et al. (eds) Big Data and Social Science. Data Science Methods and Tools for Research and Practice, 2nd ed. Business Expert Press, 2021.
- [Sha22] Shan J. et al. SQL for Data Analytics: Harness the Power of SQL to Extract Insights from Data. Packt, 2022.
- [Kha22] Khan S. Big Data and Analytics. Notion Press, 2022.
- [Per20] Perrin J.-G. Spark in Action: Covers Apache Spark 3 with Examples in Java, Python, and Scala, 2nd ed. Manning, 2020.
- [Kni22] Knight D. et al. Microsoft Power BI Quick Start Guide: The Ultimate Beginner's Guide to Data Modeling, Visualization, Digital Storytelling, and More. Packt, 2022.
- [Mil22] Milligan J.N. et al. Learning Tableau 2022: Create Effective Data Visualizations, Build Interactive Visual Analytics, and Improve Your Data Storytelling Capabilities. Packt, 2022.
- [Vit18] Vitantonio J.I., Mahler M. Mastering Qlik Sense: Expert Techniques on Self-service Data Analytics to Create Enterprise Ready Business Intelligence Solutions. Packt, 2018.
- [Ram23] Ramjan S., Sunkpho J. Principles and Theories of Data Mining with RapidMiner. IGI Global, 2023. DOI: [10.4018/978-1-6684-4730-7](#).
- [Gre20] Greco C. Data Science Tools. R, Excel, KNIME, and OpenOffice. Mercury Learning and Information, 2020. DOI: [10.1515/9781683925811](#).
- [For23] Fortino A. Data Mining and Predictive Analytics for Business Decisions. A Case Study Approach. Mercury Learning and Information, 2023. DOI: [10.1515/9781683926740](#).
- [Lee22a] Lee J., Lee C.-F. Essentials of Excel VBA, Python, and R. Volume I: Financial Statistics and Portfolio Analysis, 2nd ed. Springer, 2022. DOI: [10.1007/978-3-031-14236-9](#).
- [Lee22b] Lee J. et al. Essentials of Excel VBA, Python, and R. Volume II: Financial Derivatives, Risk Management and Machine Learning, 2nd ed. Springer, 2022. DOI: [10.1007/978-3-031-14283-3](#).
- [Mou21] Mount G. Advancing into Analytics. From Excel to Python and R. O'Reilly, 2021.
- [Sah21] Sahay A. Essentials of Data Science and Analytics. Statistical Tools, Machine Learning, and R-Statistical Software Overview. Business Expert Press, 2021. DOI: [10.4128/9781631573460](#).
- [Sco23] Scott-Boyer M.P. et al., "Use of Elasticsearch-Based Business Intelligence Tools for Integration and Visualization of Biological Data" // Briefings in Bioinformatics. 2023. Vol. 24, No. 6, pp. bbad348. DOI: [10.1093/bib/bbad348](#). EDN: [KFJZCW](#).
- [Li19] Li Y. et al. , "A cloud-based framework for large-scale log mining through Apache Spark and Elasticsearch" // Applied Science. 2019. Vol. 9, no. 1114, pp. 109-127. DOI: [10.3390/app9061114](#).
- [Cho18] Cholomskis A. et al. , "Cloud Software Performance Metrics Collection and Aggregation for Auto-Scaling Module" // Information and Software Technologies (ICIST). Communications in Computer and Information Science. 2018, No. 920, pp. 2-10. DOI: [10.1007/978-3-319-99972-2_10](#).
- [Пру23] Пруцков А.В. Способ выявления книг для начального освоения комплекса программ Elastic Stack и его результаты // Int. J. of Open Information Technologies. 2023. Т. 11, № 11. С. 53-57. DOI: [10.13140/RG.2.2.17164.73601](#). EDN: [TKFHUB](#). [[Prutzkow A. V. "An approach to identify books for the initial learning of the elastic stack and its results" // Int. J. of Open Information Technologies, 2023. Vol. 11, No. 11, pp. 53-57. DOI: [10.13140/RG.2.2.17164.73601](#).. (In Russian).]]
- [Pru25] Prutzkow A., "Elasticsearch in Medicine: A Technical Survey" // Information Society. 2025. No. 4, pp.31-38.

Поступила в редакцию 22 апреля 2025 г.

МЕТАДАННЫЕ / METADATA

Title: Analysis of university staff participation in scientific events using Elasticsearch.

Abstract: Data analysis is the basis for making management decisions. There are various programs to automate data analysis. One of the types of data analysis is aggregation. The most diverse methods of data aggregation compared to other programs are provided by the Elasticsearch (ES) program from the Elastic Stack. ES is a search engine, one of the functions of which is data aggregation. The purpose of the study is to demonstrate the capabilities of ES in aggregating data on plans for participation of university staff in scientific events. Using the term, cardinality, value count, and histogram aggregations, the following were identified: 1) university departments with the largest number of event participations; 2) events with the largest number of request in one of the departments; 3) events with the largest number of requests, held at the university and outside the university; 4) distribution of the staff depending on the number of planned participation in events. Based on the results of the aggregations, the author made conclusions of the uneven distribution of requests across departments and the preference for participation in events organized by the university, and others. Aggregation queries have explanations. The conducted data aggregation shows underestimation of ES in solving this problem.

Key words: Elasticsearch; Elastic Stack; data analysis; data aggregation; aggregation queries; staff; university; conference.

Cite: Prutzkow A. V. "Analysis of university staff participation in scientific events using Elasticsearch" // SIIT. 2025. Vol. 7, no. 4 (23), pp. 38–48. EDN [JKRJOA](#).

Language: Russian.

Об авторе / About the author**ПРУЦКОВ Александр Викторович**

Рязанский государственный радиотехнический университет имени В. Ф. Уткина (РГРТУ),

Липецкий государственный педагогический университет имени П. П. Семёнова-Тян-Шанского (ЛГПУ),
Россия.

Проф. каф. вычислительной и прикладной математики РГРТУ, каф. информатики, информационных технологий и защиты информации ЛГПУ. Д-р. техн. наук (Рязанск. гос. радиотехн. ун-т, 2015). Иссл. в обл. программирования и автоматической обработки текстов.

E-mail: mail@prutzkow.com

ORCID: <http://orcid.org/0000-0002-4110-5269>

PRUTZKOW Alexander Viktorovich

Ryazan State Radiotechnical University named after V. F. Utkin (RGRTU),

Lipetsk State Pedagogical University named after P. P. Semenov-Tyan-Shansky (LSPU),
Russia.

Prof. of the Dept of Computational and Applied Mathematics of RGRTU and Dept of Computer Science, Information Technology and Information Security of LGPU. Doctor of Engineering Sciences (Ryazan State Radio Engineering Univ., 2015). Research in the field of programming and natural language processing.

E-mail: mail@prutzkow.com

ORCID: <http://orcid.org/0000-0002-4110-5269>