Способ выявления книг для начального освоения комплекса программ Elastic Stack и его результаты

А.В. Пруцков

***Аннотация*—При подготовке учебного курса по информационно-поисковым системам и написании учебного пособия к нему возникла проблема выбора книг, посвященных комплексу программ Elastic Stack, для его начального освоения этого комплекса. Были найдены 12 книг по этой тематике. Предложен способ выявления книг для начального освоения комплекса программ Elastic Stack. Способ включает следующие этапы: 1) разработку способа оценки пригодности книг для начального освоения комплекса программ Elastic Stack; 2) оценку полноту изложения разделов с помощью экспертов; 3) дополнительное исследование книг с помощью вычисленных значений суммарной полноты разделов и кластеризации; 4) формулирование вывода о книгах, наиболее подходящих для начального освоения этого комплекса. Разработка способа оценки пригодности книг для начального освоения комплекса программ Elastic Stack состоит из следующих шагов: 1) выделение разделов, необходимых для освоения комплекса программ Elastic Stack; 2) разработку шкалы оценки полноты; 3) определение значений полноты изложения разделов книг по разработанной шкале. Для книг выделены следующие разделы и определено их содержание: поиск, агрегации, усовершенствование поиска, сценарии выполнения, внутреннее устройство Elasticsearch, индексы и документы, Logstash, Kibana. Автором статьи и привлеченными экспертами были определены значения полноты разделов. Проведено вычисление суммарной полноты разделов и кластеризация на три группы. На основе результатов этих действий сделан вывод о том, что наиболее подходящими для начального освоения комплекса Elastic Stack являются книги авторов Р. Георге и др.; Г. Гормли, З. Тонга; М. Конды; А. Паро.**

***Ключевые слова*—Информационно-поисковые системы, комплекс программ Elastic Stack, система Elasticsearch, система Kibana, система Logstash.**

I. Введение

Информационно-поисковые системы посещаются пользователями практически каждый сеанс работы в сети Интернет. Кроме информационно-поисковых систем для конечных пользователей, таких как Яндекс, Google, Duckduckgo, Baidu, существуют информационно-поисковые системы как сервис, например, комплекс программ Elastic Stack [1]. Второй тип информационно-поисковых систем может использоваться как самостоятельная информационная система или как часть более крупной информационной системы. Применение информационно-поисковых систем для решения большого числа практических и научных задач (некоторые из них перечислены в библиографическом списке статьи [2]) привело к необходимости их преподавания в вузах.

При написании учебного пособия [3-4] для курса по информационно-поисковым системам и комплексу программ Elastic Stack потребовалась справочная литература для решения следующих задач:

- познакомиться с устройством, основными возможностями и принципами работы комплекса программ Elastic Stack;
- выявить используемую терминологию;
- проанализировать последовательность изложения материала;
- изучить более глубоко по сравнению с документацией разработчика аспекты работы с комплексом программ Elastic Stack.

Комплексу программ Elastic Stack посвящено не так много книг, выпущенных в последние годы, среди которых [5-16]. Эти книги, в основном на английском языке, были использованы при написании учебного пособия. Книги [17-18] посвящены комплексу программ Elastic Stack, но не предназначены для начального освоения этого комплекса.

Книги имеют разное назначение и в разной степени подходят для решения перечисленных выше задач. Освоение комплекса программ Elastic Stack требуется не только автору статьи, но и другим преподавателям и работникам отрасли информационных технологий. Поэтому необходимо понимать, какие книги подходят для начального освоения этого комплекса.

Выбор книги для начального освоения навыка важен, так как эта книга определит взгляд на навык до полного его освоения.

II. Обзоры книг о комплексе программ Elastic Stack

Обзор книг о комплексе программ Elastic Stack сделан в [19]. Для каждой книги дана краткая аннотация и перечислены основные понятия, которая она содержит.

Статья получена 6 сентября 2023 г.

Пруцков Александр Викторович, Рязанский государственный радиотехнический университет имени В. Ф. Уткина, 390005, Российская Федерация, Рязань, ул. Гагарина, 59/1; Рязанский государственный медицинский университет имени академика И. П. Павлова, 390026, Российская Федерация, Рязань, ул. Высоковольтная, 9; Липецкий государственный педагогический университет имени П. П. Семенова-Тян-Шанского, 398020, Российская Федерация, Липецк, ул. Ленина, 42 (e-mail: mail@prutzkow.com).

Несмотря на то, что обзор сделан в 2023 г., многие книги последних лет в нем отсутствуют. Некоторые книги представлены своими непоследними изданиями. Другим обзором книг по этой тематике является обзор в [20]. Обзор также не рассматривает книги последних лет изданий. Для каждой книги указано количество страниц и приведена краткая аннотация.

III. Цель работы

Целью работы является выявление книг, посвященных комплексу программ Elastic Stack и подходящих для его начального освоения.

IV. Способ выявления книг для начального освоения комплекса программ Elastic Stack

Предлагается способ выявления книг для начального освоения комплекса программ Elastic Stack. Способ включает выполнение следующих этапов:

1) разработать способ оценки пригодности книг для начального освоения комплекса программ Elastic Stack;
2) используя способ из п. 1, оценить полноту изложения разделов с помощью экспертов;
3) дополнительно исследовать книги с помощью суммарной полноты разделов (суммы взвешенных значений полноты) и кластеризации;
4) сделать вывод о книгах, наиболее подходящих для начального освоения этого комплекса.

Разработка способа оценки пригодности книг для начального освоения комплекса программ Elastic Stack состоит из следующих шагов:

1) выделить разделы, необходимые для освоения комплекса программ Elastic Stack;
2) разработать шкалу оценки полноты;
3) определить значения полноты изложения разделов книг по разработанной шкале.

Было решено охарактеризовать книги субъективными и объективными параметрами. К субъективным параметрам относятся полнота изложения разделов для изучения комплекса программ Elastic Stack. К объективным параметрам относятся метаданные о книгах: количество страниц, год издания.

V. Разделы, необходимые для освоения комплекса программ Elastic Stack

Для каждого раздела были определены значимость (вес) и содержание, требуемое для начального освоения комплекса программ Elastic Stack (таблица 1).

Таблица 1 – Разделы, их значимость и содержание

Раздел	Значимость	Содержание
Поиск	3	Команда search; поисковые запросы; полнотекстовый поиск и поиск по термам
Агрегации	3	Классификация агрегаций; виды агрегаций
Усовершенствование поиска	2	Команды termvectors, explain, analyze, highlight; автозавершение
Сценарии выполнения	1	Script; runtime fields; scripted fields
Внутреннее устройство Elasticsearch	2	Кластеры; узлы; типы узлов; сегменты; выборы управляющего узла
Индексы и документы	3	Понятие индекса; создание, удаление индексов; структура документов; типы данных; создание, удаление, обновление документов
Logstash	2	Конфигурации; входные расширения; расширения фильтров; выходные расширения
Kibana	1	Назначение и принципы работы; отправка запросов в систему Elasticsearch; другие возможности

Комментарии к разделам и их содержанию.

- Перечисленное содержание разделов нестрогое. Если отсутствие одного из пунктов содержания компенсировалось более полным изложением других пунктов, то считалось, что раздел описан полно.
- Раздел «Сценарии выполнения» является необязательным для начального освоения комплекса программ Elastic Stack. Его наличие углубляет понимание работы с этим комплексом.

VI. Шкала оценивания

Для оценивания полноты раздела была введена следующая шкала:

1) не описан;
2) описан кратко;
3) описан скорей полно, чем кратко;
4) описан полно.

Под полным понимается описание, позволяющее применять понятия на практике для решения задач начального и среднего уровня.

Такая шкала выбрана для дальнейшей кластеризации книг.

VII. Значения параметров книг

Значения полноты изложения разделов книг оценивал автор статьи и три привлеченных эксперта. Эксперты обладают разной степенью освоения комплекса программ Elastic Stack. Полученные от экспертов оценки значений полноты по предложенной шкале усреднялись и округлялись к ближайшему целому.

Книги получили следующие значения субъективных и объективных параметров (таблица 2).

Книга [13] является справочником. В ней изложен порядок выполнения операций, но понятия объяснены кратко. Поэтому эта книга не получила высших оценок.

Таблица 2 – Значения параметров книг

Раздел	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]
Поиск	4	2	4	2	2	4	4	4	3	2	4	3
Агрегации	4	1	4	2	2	4	4	4	3	1	4	4
Усовершенствование поиска	2	2	3	2	1	4	2	4	3	1	3	3
Сценарии выполнения	2	1	3	2	2	3	2	3	3	1	1	2
Внутреннее устройство Elasticsearch	3	3	3	4	2	4	4	4	3	2	3	3
Индексы и документы	4	3	4	3	2	4	4	4	3	2	2	4
Logstash	4	1	1	4	4	1	1	2	1	4	3	3
Kibana	4	1	4	4	3	1	1	3	4	3	4	4
Количество страниц	519	176	396	474	206	498	719	475	750	180	781	538
Год издания	2019	2015	2017	2022	2015	2016	2015	2023	2022	2016	2021	2019

VIII. Суммарная полнота изложения разделов книг

Для каждой книги была вычислена суммарная полнота изложения разделов – сумма произведений значений полноты изложения разделов на значимость раздела (таблица 3).

Таблица 3 – Книги, отсортированные по суммарной полноте изложения разделов

Книга	Суммарная полнота изложения разделов
[12]	62
[5]	60
[10]	58
[7]	57
[16]	57
[11]	53
[15]	53
[13]	48
[8]	47
[9]	37
[14]	33
[6]	32

IX. Кластеризация книг

Книги, подходящие для начального освоения комплекса программ Elastic Stack, были разбиты на четыре группы кластеризацией методом k-средних по значениям полноты изложения разделов с учетом следующего:

- число кластеров должно быть небольшим, чтобы книги, подходящие для начального освоения комплекса программ Elastic Stack, оказались в одной группе;
- значений суммы квадратов расстояний от объектов до центройдов для различного числа кластеров (таблица 4).

Группы составили следующие книги (таблица 5).

Книги были кластеризованы в среде RGui с помощью языка R.

Таблица 4 – Значения суммы квадратов расстояний от объектов до центройдов для различного числа кластеров

Число кластеров	Сумма квадратов расстояний от объектов до центройдов
1	490591
2	190253
3	14955,5
4	7590
5	3216,5
6	2811
7	2792,5

Таблица 5 – Группы книг, подходящих для начального освоения комплекса программ Elastic Stack, полученные кластеризацией

Группа	Книги
1	[7, 10-13]
2	[8-9, 14]
3	[5, 15-16]
4	[6]

X. Выявление книг, подходящих для начального освоения комплекса программ Elastic Stack

С учетом следующих данных:

- значений параметров книг;
- суммарной полноты изложения разделов книг;
- результатов кластеризации книг,

экспертами был сделан вывод, что наиболее подходящими для начального освоения комплекса программ Elastic Stack являются книги [10-13] из группы 1 (таблица 5).

При выводе учитывалось то, что на освоение нового материала влияет не только его полнота, но и стиль изложения, наличие примеров и пояснений к ним, другие субъективные причины.

XI. Заключение

Были получены следующие результаты.

1. Поставлена задача выявления книг, подходящих для начального освоения комплекса программ Elastic Stack. Эта задача возникла в связи с необходимостью подготовки учебного курса, посвященного этому комплексу.
2. Для решения этой задачи предложен способ выявления книг для начального освоения комплекса программ Elastic Stack. Этот способ состоит в разработке способа оценки пригодности книг для начального освоения комплекса программ Elastic Stack, оценке полноты изложения разделов экспертами, исследовании книг с помощью суммарной полноты разделов (суммы взвешенных значений полноты) и кластеризации, формулировании вывода о книгах, наиболее подходящих для начального освоения этого комплекса.
3. Выделены следующие разделы, необходимые для освоения комплекса программ Elastic Stack: поиск, агрегации, усовершенствование поиска, сценарии выполнения, внутреннее устройство Elasticsearch,

индексы и документы, Logstash, Kibana. Для каждого раздела определено его содержание и значимость. Для оценки полноты раздела предложена шкала из четырех значений: не описан; описан кратко; описан скорей полно, чем кратко; описан полно.

4. Получены экспертные оценки значений полноты изложения разделов, необходимых для начального освоения комплекса программ Elastic Stack. Экспертами выступили автор и три специалиста из сферы информационных технологий, работающие с этим комплексом.
5. Дополнительно получены суммы взвешенных значений полноты изложения разделов, книги кластеризованы по их значениям параметров на три группы.
6. На основе результатов, полученных в пп. 4-5, экспертами сделан вывод, что наиболее подходящими для начального освоения комплекса программ Elastic Stack являются книги [10-13].

Оценка содержания и результаты кластеризации могут быть использованы исследователями и программистами при освоении комплекса программ Elastic Stack.

Предложенный способ выявления книг для начального освоения комплекса программ Elastic Stack может быть применен для выделения объектов с требуемыми свойствами из множества объектов.

Для расширения знаний о комплексе программ Elastic Stack была организована небольшая секция Всероссийской научно-практической конференции с международным участием «Информационный обмен в междисциплинарных исследованиях II» (см. например [21]).

Библиография

[1] Free and Open Search: The Creators of Elasticsearch, ELK & Kibana | Elastic – URL: https://www.elastic.co – Дата посещения 25.03.2022.

[2] Пруцков А.В. Способ поиска адресов с неполным текстом запроса в системе Elasticsearch // Информационный обмен в междисциплинарных исследованиях II: сб. тр. Всерос. науч.-практ. конф. с междунар. участием. – Рязань: Академия ФСИН России, 2023. – С. 246-250.

[3] Пруцков А.В. Информационно-поисковая система Elasticsearch: учеб. пособие: в 2 т. – Рязань: РГРТУ, 2023. – Т. 1. – 172 с.

[4] Пруцков А.В. Информационно-поисковая система Elasticsearch: учеб. пособие: в 2 т. – Рязань: РГРТУ, 2023. – Т. 2. – 184 с.

[5] Шукла П., Кумар Ш. Elasticsearch, Kibana, Logstash и поисковые системы нового поколения: пер. с англ. – СПб.: Питер, 2019. – 519 с.

[6] Akdoğan H. Elasticsearch Indexing. Packt, 2015.

[7] Andhavarapu A. Learning Elasticsearch. Distributed Real-Time Search and Analytics with Elasticsearch 5.x. Packt, 2017.

[8] Athick A. Getting Started with Elastic Stack 8.0. Packt, 2022.

[9] Chhajed S. Learning ELK Stack. Packt, 2015.

[10] Gheorghe R., Hinman M.L., Russo R. Elasticsearch in Action. Manning, 2016.

[11] Gormley G., Tong Z. Elasticsearch: The Definitive Guide. O'Reilly, 2015.

[12] Konda M. Elasticsearch in Action, 2nd ed. Manning, 2023.

[13] Paro A. Elasticsearch 8.x Cookbook, 5th ed. Packt, 2022.

[14] Sharma V. Beginning Elastic Stack. Apress, 2016.

[15] Srivastava A. Learning Elasticsearch 7.x. Index, Analyze, Search and Aggregate Your Data Using Elasticsearch. BPB, 2021.

[16] Wong W.T. Advanced Elasticsearch 7.0. A Practical Guide to Designing, Indexing, and Querying Advanced Distributed Search Engines. Packt, 2019.

[17] Кольер Р., Монтонен К., Азарми Б. Машинное обучение в Elastic Stack: пер. с англ. – М.: ДМК Пресс, 2021. – 380 с.

[18] Пиз Э. Активное выявление угроз с Elastic Stack. Построение надежного стека безопасности: предотвращение, обнаружение и оповещение: пер. с англ. – М.: ДМК Пресс, 2022. – 326 с.

[19] Best Books to Learn Elasticsearch and Kibana in 2023 | ComputingForGeeks, 2023 – URL: https://computingforgeeks.com/best-books-to-learn-elasticsearch-and-kibana/ – Дата посещения 26.02.2023.

[20] Best Elasticsearch Books From Beginner To Expert – URL: https://whatpixel.com/best-elasticsearch-books/ – Дата посещения 26.02.2023.

[21] Белов В.А. Опыт использования Elasticsearch в качестве центрального хранилища при разработке платформы HR-аналитики // Информационный обмен в междисциплинарных исследованиях II: сб. тр. Всерос. науч.-практ. конф. с междунар. участием. – Рязань: Академия ФСИН России, 2023. – С. 236-245.

An approach to identify books for the initial learning of the Elastic Stack and its results

Alexander Prutzkow

***Abstract*—While we prepared an educational course on information retrieval and the search engines, wrote a manual to it, the problem of choosing books of the Elastic Stack for the its initial learning arose. We found 12 books on this topic. We introduce an approach to identify books for the initial learning of the Elastic Stack. The method includes the following steps: (1) development of a way to evaluate the suitability of books for the initial learning of the Elastic Stack; (2) evaluating of the completeness of the content of sections using experts; (3) an additional study of books using the calculated values of the total completeness of sections and clustering; (4) formulation of the conclusion about books that are most suitable for the initial learning of the Elastic Stack. Development of the way for evaluating the suitability of books for the initial learning of the Elastic Stack consists of the following steps: (1) selection of the sections necessary for learning of the Elastic Stack; (2) the development of a completeness assessment scale; (3) determination of the values of parameters of completeness of the exposition of sections of books on the developed scale. We select the following sections for books and determine their content: search, aggregation, improvement of search, scripts, internal structure of Elasticsearch, indexes and documents, Logstash, and Kibana. The author of the article and attracted experts determined the values of the completeness of the sections. We calculate of the total completeness of sections and cluster into three groups. Based on the results of these actions, we conclude the books of the authors of R. Gheorghe et al.; G. Gormley, Z. Tonga; M. Konda; A. Paro are suitable for the initial learning of the Elastic Stack.**

***Keywords*—Search engines, Elastic Stack, Elasticsearch engine, Kibana system, Logstash system.**

REFERENCES

[1] Free and Open Search: The Creators of Elasticsearch, ELK & Kibana | Elastic – URL: https://www.elastic.co – Last accessed: 25.03.2022.

[2] Prutzkow A.V. Sposob poiska adresov s nepolnym tekstom zaprosa v sisteme Elasticsearch [Approach to the Search for Addresses with Incomplete Query Text in the Elasticsearch Engine] // Informatsionnyj obmen v mezhdistsiplinarnykh issledovanijakh II: sb. tr. Vseros. nauch.-prakt. konf. s mezhdunar. uchastiem. – Rjazan': Akademija FSIN Rossii, 2023. – S. 246-250 [in Rus].

[3] Prutzkow A.V. Informatsionno-poiskovaja sistema Elasticsearch [Elasticsearch Engine]: ucheb. posobie: v 2 t. – Rjazan': RGRTU, 2023. – T. 1. – 172 s. [in Rus].

[4] Prutzkow A.V. Informatsionno-poiskovaja sistema Elasticsearch [Elasticsearch Engine]: ucheb. posobie: v 2 t. – Rjazan': RGRTU, 2023. – T. 2. – 184 s. [in Rus].

[5] Shukla P., Kumar S. Distributed search, analytics, and visualization using Elasticsearch, Logstash, Beats, and Kibana, 2nd ed. Packt, 2019.

[6] Akdoğan H. Elasticsearch Indexing. Packt, 2015.

[7] Andhavarapu A. Learning Elasticsearch. Distributed Real-Time Search and Analytics with Elasticsearch 5.x. Packt, 2017.

[8] Athick A. Getting Started with Elastic Stack 8.0. Packt, 2022.

[9] Chhajed S. Learning ELK Stack. Packt, 2015.

[10] Gheorghe R., Hinman M.L., Russo R. Elasticsearch in Action. Manning, 2016.

[11] Gormley G., Tong Z. Elasticsearch: The Definitive Guide. O'Reilly, 2015.

[12] Konda M. Elasticsearch in Action, 2nd ed. Manning, 2023.

[13] Paro A. Elasticsearch 8.x Cookbook, 5th ed. Packt, 2022.

[14] Sharma V. Beginning Elastic Stack. Apress, 2016.

[15] Srivastava A. Learning Elasticsearch 7.x. Index, Analyze, Search and Aggregate Your Data Using Elasticsearch. BPB, 2021.

[16] Wong W.T. Advanced Elasticsearch 7.0. A Practical Guide to Designing, Indexing, and Querying Advanced Distributed Search Engines. Packt, 2019.

[17] Collier R., Montonen C., Azarmi B. Machine Learning with the Elastic Stack, 2nd ed. Packt, 2021.

[18] Pease A. Threat Hunting with Elastic Stack. Packt, 2021.

[19] Best Books to Learn Elasticsearch and Kibana in 2023 | ComputingForGeeks, 2023 – URL: https://computingforgeeks.com/best-books-to-learn-elasticsearch-and-kibana/ – Accessed 26.02.2023.

[20] Best Elasticsearch Books From Beginner To Expert – URL: https://whatpixel.com/best-elasticsearch-books/ – Last accessed 26.02.2023.

[21] Belov V.A. Opyt ispolzovanija Elasticsearch v kachestve tsentralnogo khranilischa pro razrabotke platformy HR-analitiki [Experience of Using Elasticsearch as a Central Storage in the Development of the HR Analytics Platform] // Informatsionnyj obmen v mezhdistsiplinarnykh issledovanijakh II: sb. tr. Vseros. nauch.-prakt. konf. s mezhdunar. uchastiem. – Rjazan': Akademija FSIN Rossii, 2023. – S. 236-245 [in Rus].