

УКД 004.6

ТЕХНОЛОГИЯ BIG DATA И СИСТЕМЫ ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА

БУРЕШКИН Никита Михайлович

магистрант

ФГБОУ ВО «МИРЭА – Российский технологический университет»

г. Москва, Россия

В статье рассматриваются системы электронного документооборота включающие в себя технологию Big Data. Электронный документооборот представляет из себя автоматизированный классический документооборот, представленный в электронном виде. Система электронного документооборота создается для повышения эффективности системы управления. Тенденция автоматизации документооборота привела к модернизации многих направлений деятельности. В настоящее время на рынке информационных технологий широкое распространение получили продукты электронного документооборота, спрос на которые вызван высокой конкурентоспособностью в данной области.

Ключевые слова: технология Big Data, электронный документооборот, классификация данных, работа с большими данными.

Электронный документооборот включает в себя большие массивы информации, которые невозможно обработать или проанализировать при помощи традиционных методов с использованием человеческого труда

и настольных компьютеров. Для решения данной проблемы была рассмотрена технология Big Data.

Big Data обладает следующими характеристиками (рисунок 1) [1].

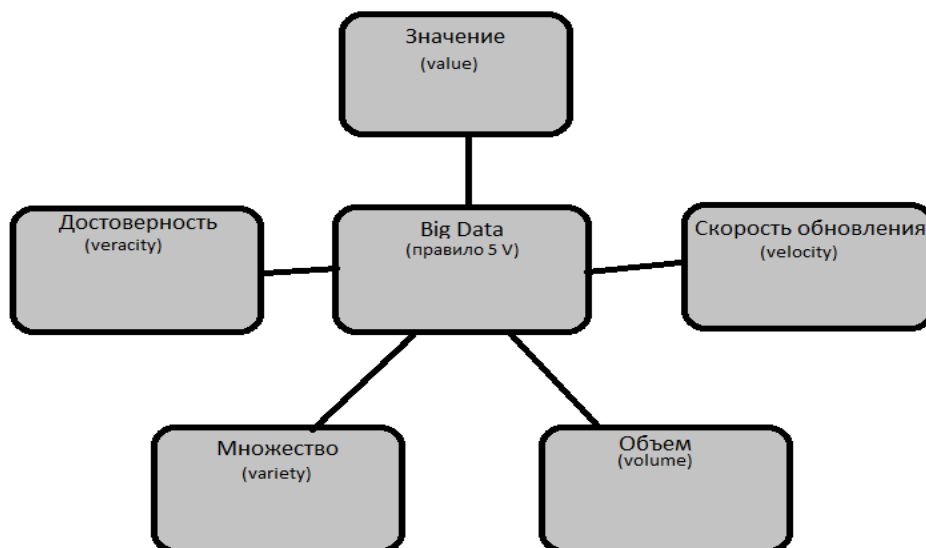


Рисунок 1. Характеристики Big Data

Значение (value) – заключается в разной сложности и переработки имеющийся информации, что затрудняет интеллектуальные системы.

Скорость обновления (velocity) – включает в себя скорость обновления и обработки данных.

Объем (volume) – относится объем свыше 150 Гб в сутки.

Множество (variety) – огромное множество видов информации, например: текстовой, видео, аудио форматы.

Достоверность (veracity) – проверка точности информации, ведь если такое количество информации будет недостоверно, анализ будет бесполезен.

Классификация данных

Данные разделяются на типы, такие как:

– структурированные;

– неструктурированные;

– естественный язык;

– машинные;

– потоковые.

Структурированные данные. Данные, которые упорядочены и хранятся в структуре, где запись каждого значения определена набором атрибутов [4]. Основным примером таких данных являются таблицы (рисунок 2) и различные форматы передачи данных: XML, JSON, YAML, CSV.

| | A | B | C | D |
|---|--------------------|--------------|-------------|--------------|
| 1 | Продукт | Кв. 1 | Кв. 2 | Общий итог |
| 2 | Шоколад | 7 446,00 ₽ | 1 625,60 ₽ | 9 071,60 ₽ |
| 3 | Мармелад | 50 796,00 ₽ | 12 492,00 ₽ | 63 288,00 ₽ |
| 4 | Багет | 12 675,00 ₽ | 10 625,00 ₽ | 23 300,00 ₽ |
| 5 | Булочки | 14 180,00 ₽ | 7 560,00 ₽ | 21 740,00 ₽ |
| 6 | Сахарный пирог | 47 280,00 ₽ | 45 479,20 ₽ | 92 759,20 ₽ |
| 7 | Шоколадное печенье | 9 438,90 ₽ | 3 496,00 ₽ | 12 934,90 ₽ |
| 8 | Всего | 141 815,90 ₽ | 81 277,80 ₽ | 223 093,70 ₽ |

Рисунок 2. Пример структурированных данных

Структурные данные удобны для человека и машины. Чаще всего данные хранятся в неструктурированном формате.

Неструктурированные данные. Данные, которые сложно подогнать под определенную модель данных, потому что содержимое

имеет переменный характер. Ярким примером неструктурированных данных, является электронное письмо. Хотя и сам формат письма определен, тем не менее поле «Содержания», содержит неструктурированные текстовые данные (рисунок 3).

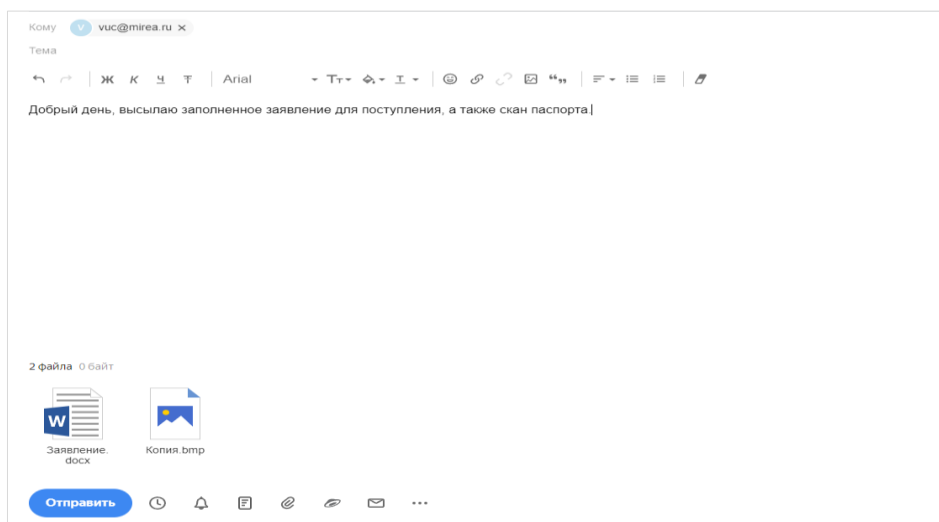


Рисунок 3. Пример неструктурированных данных

Естественный язык. Данные, полученные на естественном языке для человека. Это может быть письменный текст и устная речь. Обработка таких данных сложна, так как требует огромное количество знаний.

Машинные данные. Машинные данные включают в себя информацию, которая автоматически генерируется любым устройством без вторжения человека. Одним из примером машинных данных является детализация звонков (рисунок 4).

Страница 1 из 10

Состоявшиеся разговоры
за период с 19.01.2012 по 19.02.2012

Сетевой ресурс: **7912 456789** Номер SIM-карты: **897010117000014561**

| Дата | Время | GMT* | Номер | Зона ПС | Зона ВТК | Услуга | Длит. мин:сек | Стоимость руб. | Изм. баланса |
|------------|----------|--------|-----------------------------|---------|----------|--------|---------------|----------------|---------------------|
| 19.01.2012 | 7:31:25 | +04:00 | <--79031234567 | | | Телеф. | 2:53 | 0,0000 | |
| 19.01.2012 | 8:23:10 | +04:00 | 79120123456 | | | Телеф. | 0:26 | 5,5966 | 19.01.2012 8:23:36 |
| 19.01.2012 | 9:34:01 | +04:00 | <--79031234567 | | | Телеф. | 5:34 | 0,0000 | |
| 19.01.2012 | 10:28:12 | +04:00 | <--Ya na_svyazi:79120123456 | | | sms i | 1 | 0,0000 | |
| 19.01.2012 | 12:44:38 | +04:00 | <--79031234567 | | | sms i | 1 | 0,0000 | |
| 19.01.2012 | 14:15:46 | +04:00 | 79120123456 | | | Телеф. | 0:51 | 1,8655 | 19.01.2012 14:16:37 |
| 19.01.2012 | 15:19:50 | +04:00 | 79120123456 | | | Телеф. | 0:30 | 1,8655 | 19.01.2012 15:20:20 |
| 19.01.2012 | 15:44:34 | +04:00 | <--sms:900 | | | sms i | 1 | 0,0000 | |
| 19.01.2012 | 16:08:42 | +04:00 | <--79031234567 | | | Телеф. | 1:30 | 0,0000 | |
| 19.01.2012 | 16:28:46 | +04:00 | <--79031234567 | | | sms i | 1 | 0,0000 | |
| 19.01.2012 | 16:28:53 | +04:00 | <--79031234567 | | | sms i | 1 | 0,0000 | |
| 19.01.2012 | 16:28:59 | +04:00 | <--79123456789 | | | sms i | 1 | 0,0000 | |
| 19.01.2012 | 16:35:50 | +04:00 | 79120123456 | | | sms o | 1 | 1,6525 | 19.01.2012 16:35:50 |
| 19.01.2012 | 16:36:41 | +04:00 | 0890 | | | Телеф. | 1:43 | 0,0000 | 19.01.2012 16:38:24 |
| 19.01.2012 | 16:44:53 | +04:00 | 79120123456 | | | Телеф. | 1:05 | 3,3005 | 19.01.2012 16:45:58 |

Рисунок 4. Пример машинных данных

Потоковые данные. В состав таких данных входят разные виды данных, к примеру: действия в играх, новости из социальных сетей, телеметрические данные, данные сформированные с помощью интернет-приложений. У потоковых данных есть одно главное отличие от остальных видов данных. Информация поступает в систему, при каком-то событии, а не загружает данные в хранилище заранее.

Работа с большими данными:

В работе с большими данными выделяют несколько этапов:

- получение информации из различных ресурсов;
- занесение информации в базу данных;
- анализ и обработка данных;
- отображение полученных данных.



Рисунок 5. Схема работы с большими данными

На первом шаге предполагается сбор необходимой информации для дальнейшей работы с ней [2]. Помимо самостоятельного поиска, можно воспользоваться различными услугами, деятельность которых направлена на поиск информации.

Второй этап включает в себя занесение информации в базу данных для дальнейшего хранения.

Анализ и обработка данных является самым важным этапом. Так как нередко полученные данные могут содержать разные ошибки, повторения, неточности. Данный этап включает в себе следующие шаги:

- **первый шаг.** Удаление лишней и несоответствующей информации;
- **второй шаг.** Объединение информации из разных источников;
- **третий шаг.** Получение пригодной к использованию информации.

Сбор данных

Для того, чтобы подойти к процессу обработки больших данных надо завершить этап сбора данных. Данные хранятся в различных форматах и поступают из разных источников.

Выделяют два способа сбора данных:

- тестирование в сети интернет;
- использование API.

Тестирование. Данный способ является традиционным. Ярким примером этого способа является обычный тест, перед прохождением которого пользователи ставят галочку напротив поля о согласии на передачу данных.

Использование API. Способ заключающийся в себе сбор записей из наборов данных, через программный интерфейс приложения (API – Application Programming Interface). Данный метод позволяет, получить всю необходимую информацию при отсутствии прямого контакта с носителем необходимой информации.

Из проведенного анализа возможностей технологии Big Data можно сделать вывод, что данная технология применима для оптимизации существующих процессов обработки данных, позволяет существенно сократить затраты на хранение и обработку и при этом обеспечить эффективность работы с данными, что является наиболее важным в процессе разработки системы электронного документооборота.

СПИСОК ЛИТЕРАТУРЫ

1. *Величко Н.А., Митрейкин И.П.* Технология BIG DATA. Анализ рынка BIG DATA // Синергия Наук. – 2018. – № 30. – С. 937-943. – URL:https://www.elibrary.ru/download/elibrary_36735848_25881783.pdf (дата обращения: 25.03.2023).
2. *Гарипова К.Р.* Концепция Bigdata // В сборнике: Решетневские чтения. Материалы XXV Международной научно-практической конференции, посвященной памяти генерального конструктора ракетно-космических систем академика М.Ф. Решетнева. В 2-х частях; под общей редакцией Ю.Ю. Логинова. – Красноярск, 2021. – С. 287-288. – URL:https://www.elibrary.ru/download/elibrary_47711895_92105996.pdf (дата обращения: 12.05.2023).
3. *Гуськов Д.С.* Перспективы использования технологии Big Data // Актуальные проблемы авиации и космонавтики. Сборник материалов VII Международной научно-практической конференции, посвященной Дню космонавтики: в 3 томах / под общей редакцией Ю.Ю. Логинова. – Красноярск, 2021. – С. 497-499. – URL:https://elibrary.ru/download/elibrary_47575580_86077248.pdf (дата обращения: 12.06.2023)
4. Структурированные данные. – URL:<https://wiki.loginom.ru/articles/structured-data.html> (дата обращения: 08.05.2022).
5. *Сырчин В.И., Агафонова О.А.* Перспективы использования технологии Big Data в формировании официальной статистической информации // Статистика прошлого, настоящего и будущего. Материалы Межрегиональной научно-практической конференции, посвященной 100-летию образования статистической службы Горного Алтая. – Горно-Алтайск, 2022. – С. 291-294. – URL:https://elibrary.ru/download/elibrary_50227242_59571359.pdf (дата обращения: 13.07.2023)

BIG DATA TECHNOLOGY AND ELECTRONIC DOCUMENT MANAGEMENT SYSTEMS

BURESHKIN Nikita Mikhailovich
Undergraduate Student
MIREA – Russian Technological University
Moscow, Russia

In the article, system research includes document management that includes big data standards. Electronic document management is a classic electronic document management presented in electronic form. The document flow output system was created to improve the efficiency of the management system. The monetary revolution of document management has led to a tendency to change many types of activities. Currently, products that increase document flow are widespread in the information technology market, the demand for which causes low competitiveness in this area.

Keywords: Big Data technology, electronic document management, data classification, working with big data.

ПРАВО

АДВОКАТ КАК УЧАСТНИК СЛЕДСТВЕННЫХ ДЕЙСТВИЙ СО СТОРОНЫ ОБВИНЕНИЯ: ЗАКОННОСТЬ ПРОВОДИМОГО ИМ ИССЛЕДОВАНИЯ

БОЧКОВОЙ Андрей Алексеевич
бакалавр
ФГБОУ ВО «Кубанский государственный аграрный университет им. И.Т. Трубилина»
г. Краснодар, Россия

Интересным представляется исследование, проведенное адвокатом по поручению следователя, который изъял у предпринимателя вещи для установления товарного знака, товарообладателя, чтобы установить ущерб, и привлечь предпринимателя к уголовной ответственности по ст. 180 УК РФ «Незаконное использование товарного знака»? И как трактовать исследование адвоката? Кем является адвокат в данном случае? Участником следственных действий? Специалистом? Или у него нет таких полномочий, и выводы его ничтожны, незаконны. Насколько это правомерно?

Ключевые слова: следственные действия, уголовное судопроизводство, следователь, адвокат, законность, товарный знак.

При расследовании преступления по ст. 180 УК «Незаконное использование товарного знака» следователь изъял из магазинов ИП А. вещи с признаками контрафактности: сумку, очки, рюкзак, кроссовки с товарными знаками Louis Vuitton, Ralph Lauren, Ray Ban. Затем следователь обратился с запросом в адвокатское бюро, которое представляет интересы компаний – правообладателей товарных

знаков, и поставил вопросы о том, является ли продукция контрафактной, кто ее правообладатель, зарегистрированы ли товарные знаки на территории РФ, заключались ли соглашения, предоставляющие право использования товарных знаков ИП А.

При расследовании преступления по ст. 180 УК «Незаконное использование товарного знака» следователь изъял из магазинов ИП А.