



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«МИРЭА □ Российский технологический университет»

РТУ МИРЭА

Институт комплексной безопасности и специального приборостроения

Кафедра «Цифровые технологии обработки данных»

ОТЧЕТ

по практической работе

«Практическая работа №1. Слабоструктурированные форматы данных»

по дисциплине «Нереляционные системы управления базами данных»

Выполнил

Смирнов И.А.

фамилия, имя, отчество

21Б0700

шифр

БСБО-11-21

группа

Проверил

К.Т.Н., доцент

ученая степень, должность

Ильин Д.Ю.

фамилия, имя, отчество

Москва 2023г.

Цель практической работы

Цель настоящей практической работы – научиться использовать форматы слабоструктурированных данных и осуществлять манипуляции с файлами в соответствующих форматах.

Задачи практической работы

Для достижения поставленной цели необходимо решить следующие задачи:

1. Сгенерировать исходный набор данных.
2. Сохранить набор данных в формате №1 согласно варианту задания.
3. Считать данные из сохраненного файла.
4. Применить к данным преобразование согласно варианту задания.
5. Сохранить измененный набор данных в формате №2 согласно варианту задания.
6. Подготовить ответы на контрольные вопросы.
7. Составить отчет о проведенной работе.

Генерируемый набор данных должен состоять минимум из 1000 записей.

Каждая запись должна содержать как минимум следующие данные:

- Идентификатор пользователя.
- Имя пользователя.
- Массив из одного и более электронных почтовых адресов пользователя.
- Дата регистрации пользователя.
- Дата последней авторизации пользователя.
- Статус аккаунта пользователя – подтвержден или нет.
- Массив из одной и более публикаций пользователя.
 - Название.
 - Описание (многострочное).
 - Объем в страницах.
 - Категория публикации.
 - Дата публикации.
 - Массив из нуля и более отзывов.

- Идентификатор пользователя.
- Текст отзыва (многострочный).
- Дата рождения пользователя.
- Пол пользователя.

Вариант 7, 5

JSON -> DSV (разделитель – знак табуляции)

Добавить отметку пользователям с возрастом меньше 14 лет

Реализация

Генерация случайных данных 1000 профилей с помощью функций и запись в JSON файл.

```
function generateRandomDate(start, end) {
    return new Date(start.getTime() + Math.random() * (end.getTime() - start.getTime()));
}

function generateRandomString() {
    const characters = 'ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz0123456789';
    let randomString = '';

    let length = Math.floor((Math.random() * 10) + 1);

    for (let i = 0; i < length; i++) {
        const randomIndex = Math.floor(Math.random() * characters.length);
        randomString += characters.charAt(randomIndex);
    }

    return randomString;
}

function generateRandomNumber(min, max) {
    return Math.floor(Math.random() * (max - min)) + min;
}

function generateRandomData() {
    const startDate = new Date(2000, 0, 1);
    const endDate = new Date();

    const userId = generateRandomNumber(1, 100000);

    function generateRandomEmails() {
        const numEmails = generateRandomNumber(1, 4);
        const emails = [];
        for (let i = 0; i < numEmails; i++) {
            emails.push(`user${generateRandomString()}@gmail.com`);
        }
        return emails;
    }

    function generateRandomPublications() {
```

```

const numPublications = generateRandomNumber(1, 4);
const publications = [];

for (let i = 0; i < numPublications; i++) {
  let description = '';
  const descriptionLength = Math.floor((Math.random() * 10) + 50);

  for (let j = 0; j < descriptionLength; j++) {
    description += generateRandomString();

    if (j < descriptionLength - 1) {
      description += ' ';
    }
  }

  function generateRandomReviews() {
    const reviews = [];
    const reviewCount = Math.floor((Math.random() * 10));

    for (let d = 0; d < reviewCount; d++) {
      let reviewDescription = '';
      const reviewsLength = Math.floor((Math.random() * 10) + 30);

      for (let k = 0; k < reviewsLength; k++) {
        reviewDescription += generateRandomString();

        if (k < reviewsLength - 1) {
          reviewDescription += ' ';
        }
      }

      reviews.push({
        userID: generateRandomNumber(1, 100000),
        description: reviewDescription,
      });
    }

    return reviews;
  }

  publications.push({
    name: `Publication ${generateRandomString()}`,
    description: `${description}`,
    pages: generateRandomNumber(1, 100),
    category: `Category ${generateRandomNumber(1, 5)}`,
    date: generateRandomDate(startDate,
endDate).toLocaleDateString(),
    reviews: generateRandomReviews(),
  });
}

return publications;
}

return {
  userId: userId,
  userName: `User${userId}`,
  emails: generateRandomEmails(),
  registrationDate: generateRandomDate(startDate,
endDate).toLocaleDateString(),
  lastLoginDate: generateRandomDate(startDate,
endDate).toLocaleDateString(),
  accountStatus: generateRandomNumber(0, 2) === 0 ? 'Confirmed' : 'Not
Confirmed',

```

```

        publications: generateRandomPublications(),
        birthDate: generateRandomDate(new Date(1970, 0, 1),
endDate).toLocaleDateString(),
        gender: generateRandomNumber(0, 2) === 0 ? 'Male' : 'Female',
    };
}

function generateDataset(numRecords) {
    const dataset = [];
    for (let i = 0; i < numRecords; i++) {
        dataset.push(generateRandomData());
    }
    return dataset;
}

const jsonDataset = generateDataset(1000);

const fs = require('fs');

fs.writeFileSync('jsonDataset.json', JSON.stringify(jsonDataset, null, 2));

```

Листинг 1 – Генерация JSON.

Добавление отметки пользователям, возрастом меньше 14 лет.

```

const fs = require('fs');
const moment = require('moment');

fs.readFile('jsonDataset.json', 'utf8', (err, data) => {

    const usersData = JSON.parse(data);

    usersData.forEach((userData, index) => {
        const currentDate = moment();
        const userBirthDate = moment(userData.birthDate, 'DD.MM.YYYY');
        const age = currentDate.diff(userBirthDate, 'years');

        if (age < 14) {
            console.log(`Пользователь ${userData.userName} младше 14 лет.`);
            userData.birthDate += '***';
        } else {
            console.log(`Пользователь ${userData.userName} старше или равен 14
годам.`);
        }
    });

    fs.writeFileSync('jsonChangedDataset.json', JSON.stringify(usersData, null,
2));
});

```

Листинг 2 – преобразование JSON.

Генерация DSV файла.

```

const fs = require('fs');

fs.readFile('jsonChangedDataset.json', 'utf8', async (err, data) => {

    const jsonData = JSON.parse(data);

    const d3dsv = await import('d3-dsv');

```

```
const delimiter = ' ';  
  
const dsvData = d3dsv.dsvFormat(delimiter).format(jsonData);  
  
fs.writeFileSync('dsvDataset.dsv', dsvData);  
});
```

Листинг 3 – Генерация DSV.

Ответы на контрольные вопросы

1) Какие слабоструктурированные форматы данных существуют? В чём их особенности?

Текстовые файлы: способны содержать любую текстовую информацию вне зависимости от структуры данных или её отсутствия

Языки разметки: языки разметки, такие как HTML, XML, могут содержать иерархическую структуру, однако возможность включения произвольных данных и атрибутов в тегах делают их слабоструктурированными.

2) Как отображается иерархия в слабоструктурированных форматах данных?

Иерархия в слабоструктурированных форматах данных может отражаться через вложенность или именованные поля. В JSON данные могут быть вложены в объекты и массивы, что позволяет создать иерархию. В XML, иерархия отображается с помощью вложенных тегов и атрибутов.

Текстовые файлы могут использовать символы-разделители для разделения данных на разные уровни.

3) Какие недостатки имеют слабоструктурированные форматы по сравнению со структурированным представлением данных?

Слабоструктурированные форматы не имеют строгой схемы данных, что усложняет проверку данных на целостность или корректность.

Слабоструктурированные форматы часто не различают типы данных, из-за чего они могут быть мало пригодными для работы с датами или числами с плавающей точкой. Слабоструктурированные данные зачастую требуют дополнительный код для парсинга.

4. Чем отличается формат YAML от JSON?

YAML использует табуляцию в качестве главного разделителя, в отличие от скобок в JSON. В связи с этим YAML данные легче в обработке человеком «на глаз».

5. Какие библиотеки могут использоваться для манипулирования слабоструктурированными данными?

Python содержит встроенные библиотеки для работы с различными слабоструктурированными видами данных (csv, json). Помимо этого популярная библиотека Pandas даёт возможность работы с слабоструктурированными dataframes и встроенные способы их перевода в другие форматы. JavaScript: В JavaScript есть встроенная поддержка JSON. Для работы с XML можно использовать DOMParser или библиотеки, такие как xml2js. Для YAML существует библиотека js-yaml. Java: В Java можно использовать библиотеки, такие как Jackson для JSON, JAXB для XML, и SnakeYAML для YAML.

Результаты

```
1  [
2  {
3    "userId": 00682,
4    "userName": "User00682",
5    "emails": [
6      "userAkm7to@gmail.com",
7      "userLl2fvLbHDz@gmail.com"
8    ],
9    "registrationDate": "08.06.2022",
10   "lastLoginDate": "23.12.2020",
11   "accountStatus": "Not Confirmed",
12   "publications": [
13     {
14       "name": "Publication Qw05t0nTm",
15       "description": "x u801Y2 ie8DUDMj 0 paXxm5k zQNUiG DY v aoF0BVe S2KjEeg sSbHmTq NFz 8mrIDBkHUC H9mYO ftPD0t 3pTBA8D9C d8Ncnv6Zx 8rrPtKlHk 9sSi5 RSfvtb9Z9",
16       "pages": 14,
17       "category": "Category 4",
18       "date": "27.01.2006",
19       "reviews": [
20         {
21           "userId": 11652,
22           "description": "a6xJ8t qUIL62K Ro0BAS0Jq8 ff N 88HEr Equ 3bfm DPXiTh svU7AaDLIM wnpqL7q6EH ls0NnAVQa ilj5F FOCERn TaA6Zefj5 kKy 0x5CUK4LWa fKkX1xL4E F",
23         },
24         {
25           "userId": 57547,
26           "description": "yQx XCntvE ruoQn mIZJNz2 wh W 79oTh3 Jzm1 cQY1A00jn DctGA MFAUj qhN9ezLwY QxL ZspTq6PKDt BLfQE PDjxF tmsB0mqA26 wrB3ehB 2X MgmIydkSMb h",
27         },
28         {
29           "userId": 22090,
30           "description": "If 0c48Lk PXCQQ n9K x8VOW6d I9 loCdDnm NjxXDU0enT f6N1hS xM1CkurDw1 xu9 h6R3 kTe cH Nk7S wPHwQ Q2esmJHB CLH1UMmm wFCT 08YEJML Q9JI b0Cp",
31         },
32         {
33           "userId": 85994,
34           "description": "QdZkjS v7BJm89i7V H5xs YsQcsHg WXSQpNLiI l571Db Hu0 Md nxb0 twT4LMXMX s9q3CBFdTK PXbgdc 6hFnc8eKr Xy 0fnP 05Xv Pc6umJ lV SVGmtp VuAIEre",
35         }
36       ]
37     }
38   ],
39   "birthDate": "03.04.2011",
40   "gender": "Male"
41 }
```

Рис 1 - Сгенерированный JSON.

```
39      "birthDate": "03.04.2011***",
40      "gender": "Male"
```

Рис 2 – Отметка у пользователя

```
1  |userId|userName|emails|registrationDate|lastLoginDate|accountStatus|publications|birthDate|gender|
2  60682|User60682|userAkm7t6@gmail.com,userLL2fvLbHdz@gmail.com|08.06.2022|23.12.2020|"Not Confirmed"|"[object Object]"|03.04.2011***|Male|
3  52443|User52443|userNfhX94R0mZ@gmail.com,userOUqbJYJ@gmail.com|24.09.2013|01.07.2010|Confirmed|"[object Object],[object Object],[object Object]"|11.01.2004|Male|
4  78347|User78347|userfswja@gmail.com,userxAr2eV@gmail.com|04.06.2023|15.12.2016|Confirmed|"[object Object],[object Object],[object Object]"|06.12.1998|Male|
5  99131|User99131|userV0Yh7@gmail.com|19.01.2004|18.09.2003|"Not Confirmed"|"[object Object],[object Object]"|24.08.1982|Male|
6  93266|User93266|userUljJ0z@gmail.com,userKUnV802Jtb@gmail.com,userT8ztM10@gmail.com|03.05.2009|10.08.2014|"Not Confirmed"|"[object Object]"|28.10.1980|Female|
7  39252|User39252|user6@gmail.com,user2@gmail.com|02.05.2022|30.03.2000|Confirmed|"[object Object],[object Object]"|09.09.1984|Female|
8  40672|User40672|user6UjvWRS4@gmail.com|06.12.2004|22.08.2008|Confirmed|"[object Object],[object Object]"|21.07.2015***|Male|
9  65938|User65938|userV6LO0FD@gmail.com,user8q6r@gmail.com|20.11.2021|08.04.2004|"Not Confirmed"|"[object Object],[object Object]"|26.12.1999|Male|
10 94657|User94657|usercBfsJT@gmail.com|12.03.2021|12.09.2006|Confirmed|"[object Object],[object Object]"|17.11.1998|Female|
11 82127|User82127|userJ@gmail.com|06.08.2004|14.04.2007|Confirmed|"[object Object],[object Object]"|29.12.1979|Male|
12 39845|User39845|userzpg@gmail.com,useroU4vcqKco@gmail.com|29.12.2016|24.02.2005|"Not Confirmed"|"[object Object],[object Object]"|18.02.2011***|Male|
13 65135|User65135|userP322lrN1m@gmail.com,userjre@gmail.com|28.12.2008|27.12.2008|Confirmed|"[object Object]"|19.01.2005|Male|
14 37921|User37921|user53@gmail.com,userCp6xFLzzRn@gmail.com|06.10.2008|11.12.2018|"Not Confirmed"|"[object Object]"|19.07.1993|Female|
15 93182|User93182|usernQA@gmail.com|19.11.2018|31.08.2018|"Not Confirmed"|"[object Object]"|02.08.2000|Male|
16 8120|User8120|userKf67LNZR1@gmail.com|31.07.2003|20.06.2006|Confirmed|"[object Object],[object Object],[object Object]"|22.03.2020***|Female|
17 75454|User75454|userAp34rnlB@gmail.com,userc@gmail.com,useruCXly@gmail.com|21.12.2019|24.08.2000|"Not Confirmed"|"[object Object],[object Object]"|31.08.2006|Female|
18 59080|User59080|usera5tWVSP6c@gmail.com|12.03.2000|17.07.2009|Confirmed|"[object Object],[object Object]"|29.01.2017***|Male|
19 12332|User12332|userpzyTEil@gmail.com|11.09.2004|25.12.2016|"Not Confirmed"|"[object Object]"|30.04.1985|Male|
20 61060|User61060|userDLm@gmail.com|27.02.2012|26.06.2017|Confirmed|"[object Object],[object Object],[object Object]"|08.07.1999|Male|
21 31639|User31639|usero@gmail.com,userIweB5M@gmail.com,user226u1pN@gmail.com|04.01.2013|26.04.2009|"Not Confirmed"|"[object Object],[object Object]"|01.06.1996|Male|
22 92636|User92636|userwdENqIBgy@gmail.com|02.04.2000|23.11.2004|"Not Confirmed"|"[object Object],[object Object],[object Object]"|05.10.2007|Male|
23 63991|User63991|userke@gmail.com|22.06.2006|27.04.2002|"Not Confirmed"|"[object Object],[object Object],[object Object]"|31.07.1990|Male|
24 60259|User60259|user6h5p5EI8ER@gmail.com,userLU@gmail.com|22.11.2017|07.06.2001|"Not Confirmed"|"[object Object]"|16.02.2012***|Male|
25 60406|User60406|useridZd@gmail.com,userP4cujgSWt@gmail.com,user6I8R8@gmail.com|14.09.2007|24.10.2019|"Not Confirmed"|"[object Object],[object Object]"|22.05.2011***|
26 66900|User66900|userBTRZjd@gmail.com,user0NSzJJ@gmail.com,userBF@gmail.com|23.09.2019|12.03.2001|"Not Confirmed"|"[object Object],[object Object]"|29.08.2001|Female|
27 67206|User67206|userDzJfI@gmail.com,userrrhgZV8A9v4@gmail.com,user3@gmail.com|07.06.2006|13.11.2013|"Not Confirmed"|"[object Object],[object Object]"|30.06.1998|Feme
28 58878|User58878|useret@gmail.com,user3d3@gmail.com|25.08.2022|25.09.2003|Confirmed|"[object Object]"|22.07.1992|Female|
29 52558|User52558|userwetUB@gmail.com,userrrk@gmail.com|14.04.2001|27.04.2020|"Not Confirmed"|"[object Object],[object Object],[object Object]"|06.09.2012***|Male|
30 57476|User57476|userMJ0y6X@gmail.com,userKBczXek@gmail.com,userYDSUWSC@gmail.com|11.05.2003|29.08.2009|Confirmed|"[object Object],[object Object]"|30.09.1975|Female|
31 94934|User94934|userfc7mnCud00@gmail.com|30.07.2002|01.10.2015|Confirmed|"[object Object]"|01.11.1987|Female|
32 35545|User35545|usersHycVuvN1@gmail.com,userP1FOy@gmail.com,user9paNovzNPC@gmail.com|05.07.2019|06.08.2001|Confirmed|"[object Object],[object Object],[object Object]"|
33 61946|User61946|userUPYiFPDE@gmail.com,useru3@gmail.com|27.07.2003|20.07.2019|Confirmed|"[object Object],[object Object]"|16.07.1973|Female|
34 50730|User50730|userYj6nU4@gmail.com,userXQJ7HzC@gmail.com,userNDB@gmail.com|03.02.2004|20.10.2021|"Not Confirmed"|"[object Object],[object Object],[object Object]"|
35 94249|User94249|userQPxy42t@gmail.com,userQJJ048FicD@gmail.com,userXyFKsD8@gmail.com|23.06.2009|24.02.2004|Confirmed|"[object Object],[object Object]"|06.02.2020**
36 37177|User37177|userd7J@gmail.com|08.08.2003|04.02.2016|Confirmed|"[object Object],[object Object]"|03.09.2018***|Female|
37 5079|User5079|userUH@gmail.com,userjdyWICrsHA@gmail.com|30.10.2016|05.06.2014|"Not Confirmed"|"[object Object],[object Object],[object Object]"|24.04.2013***|Female|
38 44930|User44930|userV1F6K@gmail.com,useru0PeIUHBKx@gmail.com,userLcISoR68p@gmail.com|04.05.2012|06.09.2001|"Not Confirmed"|"[object Object],[object Object]"|12.03.1
39 86130|User86130|userHPM@gmail.com|10.09.2017|11.12.2014|Confirmed|"[object Object],[object Object],[object Object]"|19.02.1990|Male|
40 13880|User13880|userV6n@gmail.com|09.10.2009|24.04.2009|Confirmed|"[object Object],[object Object]"|21.02.2006|Female|
41 96873|User96873|userzmdHSmb@gmail.com|07.03.2009|10.11.2021|"Not Confirmed"|"[object Object]"|23.12.1991|Male|
42 78846|User78846|userxps@gmail.com|02.09.2020|09.01.2020|Confirmed|"[object Object],[object Object],[object Object]"|22.11.2002|Female|
```

Рис 3 - Сгенерированные данные в формате DSV