



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

КАФЕДРА КОМПЬЮТЕРНЫЕ СИСТЕМЫ И СЕТИ (ИУ6)

НАПРАВЛЕНИЕ ПОДГОТОВКИ 09.04.01 Информатика и вычислительная техника

МАГИСТЕРСКАЯ ПРОГРАММА 09.04.01/12 Интеллектуальный анализ больших
данных в системах поддержки принятия решений.

О Т Ч Е Т

по лабораторной работе № 10

Вариант № 11

Название: Spark

Дисциплина: Языки программирования для работы с большими данными

Студент

ИУ6-23М

(Группа)

(Подпись, дата)

С.В.Мельников

(И.О. Фамилия)

Преподаватель

(Подпись, дата)

П.В. Степанов

(И.О. Фамилия)

Москва, 2023

Цель работы

Изучить работу со Spark на языке программирования Java.

Задание:

- 1) Выбрать любой датасет на kaggle.com
- 2) Сделать 10 выборок данных по выбранной предметной области

В ходе лабораторной работы используется датасет с рейтингами ВУЗов (рисунок 1).

rank_o...	ra...	name (3)	scores_ov...	scores_overall...	scores_tea...	scores_teachin...	scores_res...	scores_researc...	scores_citati...	scores_citations...	scores_industry...	scores_industry_inco...	sco
rank_order	rank	name	scores_overall	scores_overall_rank	scores_teaching	scores_teaching_rank	scores_research	scores_research_rank	scores_citations	scores_citations_rank	scores_industry_income	scores_industry_income_rank	scores
1													
10	1	University of Oxford	96.4	10	92.3	5	99.7	1	99.0	28	74.9	158	96.2
20	2	Harvard University	95.2	20	94.8	1	99.0	3	99.3	20	49.5	418	89.5
30	3	University of Cambridge	94.8	30	90.9	7	99.5	2	97.0	61	54.2	335	95.8
40	4	Stanford University	94.8	40	94.2	2	96.7	7	99.8	16	65.0	232	79.8
50	5	Massachusetts Institute of Technology	94.2	50	90.7	8	93.6	11	99.8	15	90.9	62	89.3
60	6	California Institute of Technology	94.1	60	90.9	6	97.0	5	97.3	55	89.8	69	83.6
70	7	Princeton University	92.4	70	87.6	12	95.9	8	99.1	25	66.0	221	80.3
80	8	University of California, Berkeley	92.1	80	86.4	14	95.8	9	99.0	27	76.8	145	78.4
90	9	Yale University	91.4	90	92.6	3	92.7	14	97.0	66	55.0	323	70.9
100	10	Imperial College London	90.4	100	82.8	16	90.8	17	98.3	40	59.8	274	97.5
110	11	Columbia University	89.4	110	89.4	10	87.7	21	97.1	57	44.8	569	79.9
120	12	ETH Zurich	89.4	120	82.6	17	95.4	10	90.7	135	59.1	279	97.7
130	13	The University of Chicago	88.9	130	86.5	13	88.8	19	97.7	49	56.2	303	74.2
..													

Рисунок 1 – Датасет с рейтингами ВУЗов

Листинг программы:

```
public class Main {  
    public static void main(String[] args) {  
        SparkSession spark =  
SparkSession.builder().appName("TestingSpark").master("local[*]").getOrCreate();  
        Dataset<Row> data = spark.read().format("csv").option("header", "true")  
            .load("./lab10/src/main/resources/2023_rankings.csv");  
        data.createOrReplaceTempView("universities");  
        System.out.println("1st DataFrame:");  
        spark.sql("SELECT * FROM universities").show(30, false);  
        System.out.println("2nd DataFrame:");  
        spark.sql("SELECT universities.name FROM universities WHERE
```

```

universities.location = 'Russian Federation')).show(30, false);
    System.out.println("3rd DataFrame:");
    spark.sql("SELECT universities.name, universities.scores_teaching FROM
universities WHERE universities.scores_teaching > 60").show(30, false);
    System.out.println("4th DataFrame:");
    spark.sql("SELECT universities.name, universities.rank FROM universities
WHERE universities.location = 'China' AND universities.rank_order > 100").show(30,
false);
    System.out.println("5th DataFrame:");
    spark.sql("SELECT universities.name, universities.aliases FROM
universities WHERE universities.rank_order > 100").show(30, false);
    System.out.println("6th DataFrame:");
    spark.sql("SELECT COUNT(*) FROM universities WHERE
universities.scores_citations_rank > 100").show();
    System.out.println("7th DataFrame:");
    spark.sql("SELECT universities.name FROM universities WHERE
universities.scores_industry_income_rank <
scores_international_outlook_rank").show(30, false);
    System.out.println("8th DataFrame:");
    spark.sql("SELECT universities.name, universities.subjects_offered FROM
universities WHERE universities.location = 'Japan'").show(30, false);
    System.out.println("9th DataFrame:");
    spark.sql("SELECT * FROM universities ORDER BY 5").show(5, false);
    System.out.println("10th DataFrame:");
    spark.sql("SELECT universities.name, universities.rank FROM universities
ORDER BY universities.rank_order DESC LIMIT 10").show(10, false);
    spark.stop();
}
}

```

Работа программы представлена на рисунках 2-11.

rank_order	rank	name	scores_overall	scores_overall_rank	scores_teaching	scores_teaching_rank	scores_research	scores_research_rank	scores_citations	scores_citations_rank	scores_industry_inc
110	11	University of Oxford	196.4	110	192.3	15	199.7	11	199.0	128	174.9
120	12	Harvard University	195.2	120	194.8	11	199.0	13	199.3	120	149.5
130	13	University of Cambridge	194.8	130	190.9	17	199.5	12	197.0	161	154.2
140	14	Stanford University	194.8	140	194.2	12	196.7	17	199.8	116	105.0
150	15	Massachusetts Institute of Technology	194.2	150	190.7	18	193.6	111	199.8	115	190.9
160	16	California Institute of Technology	194.1	160	190.9	16	197.0	15	197.3	155	189.8
170	17	Princeton University	192.4	170	187.6	112	195.9	18	199.1	125	166.0
180	18	University of California, Berkeley	192.1	180	186.4	114	195.8	19	199.0	127	176.8
190	19	Yale University	191.4	190	192.6	13	192.7	114	197.0	166	155.0
100	10	Imperial College London	190.4	100	182.8	116	190.8	117	198.3	140	159.8
110	11	Columbia University	189.4	110	189.4	110	187.7	121	197.1	157	144.8
120	11	ETH Zurich	189.4	120	182.6	117	195.4	110	190.7	1135	159.1
130	113	The University of Chicago	188.9	130	186.5	113	188.8	119	197.7	149	156.2
140	14	University of Pennsylvania	188.8	140	186.0	115	188.8	120	197.0	165	175.8
150	15	Johns Hopkins University	188.3	150	179.4	121	191.5	115	197.0	164	189.5
160	16	Tsinghua University	188.2	160	190.1	19	197.4	14	188.0	1182	1100.0
170	17	Peking University	188.1	170	192.5	14	196.7	16	180.4	1308	191.8
180	18	University of Toronto	187.4	180	177.3	125	193.3	112	192.8	1113	165.5
190	19	National University of Singapore	187.1	190	176.4	126	193.0	113	190.2	1148	187.0
200	20	Cornell University	185.9	200	180.2	120	186.1	122	197.3	156	140.4
210	21	University of California, Los Angeles	185.8	210	180.4	119	188.9	118	195.4	186	158.8
220	22	UCL	185.7	220	174.5	129	185.4	123	197.9	146	144.5
230	23	University of Michigan-Ann Arbor	182.9	230	179.3	122	184.6	124	193.6	1103	148.7
240	24	New York University	182.7	240	174.2	130	184.0	125	195.0	193	144.6
250	25	Duke University	182.6	250	178.1	123	176.2	137	195.7	183	199.6
260	26	Northwestern University	182.1	260	171.7	132	180.7	130	197.5	152	185.0
270	27	University of Washington	182.1	270	171.6	133	182.8	127	198.9	129	153.9
280	28	Carnegie Mellon University	181.1	280	165.4	146	181.8	129	198.7	131	155.2
290	29	University of Edinburgh	179.8	290	166.9	142	174.5	144	197.1	158	140.9
300	30	Technical University of Munich	179.3	300	169.8	134	182.2	128	184.5	1241	1100.0

only showing top 30 rows

Рисунок 2 – Работа программы (запрос 1)

```

20/09/07 07:20:47 INFO Generator: Data generated at 20/07/17 ms
+-----+
|name|
+-----+
|Lomonosov Moscow State University|
|Moscow Institute of Physics and Technology (MIPT)|
|Peter the Great St Petersburg Polytechnic University|
|HSE University|
|National Research Nuclear University MEPhI|
|Bauman Moscow State Technical University|
|Saint-Petersburg Mining University|
|Financial University under the Government of the Russian Federation|
|National University of Science and Technology (MISiS)|
|Plekhanov Russian University of Economics|
|RUDN University|
|Tomsk State University|
|ITMO University|
|Kazan Federal University|
|Novosibirsk State University|
|Saint Petersburg State University|
|Sechenov University|
|South Ural State University|
|Don State Technical University|
|Tomsk Polytechnic University|
|Ufa State Petroleum Technological University|
|Altai State University|
|Far Eastern Federal University|
|Immanuel Kant Baltic Federal University|
|Irkutsk National Research Technical University|
|Kazan National Research Technical University|
|Kazan National Research Technological University|
|Kazan State Medical University|
|Lobachevsky State University of Nizhni Novgorod|
|Moscow Aviation Institute|
+-----+
only showing top 30 rows

```

Рисунок 3 – Работа программы (запрос 2)

name	scores_teaching
University of Oxford	92.3
Harvard University	94.8
University of Cambridge	90.9
Stanford University	94.2
Massachusetts Institute of Technology	90.7
California Institute of Technology	90.9
Princeton University	87.6
University of California, Berkeley	86.4
Yale University	92.6
Imperial College London	82.8
Columbia University	89.4
ETH Zurich	82.6
The University of Chicago	86.5
University of Pennsylvania	86.0
Johns Hopkins University	79.4
Tsinghua University	90.1
Peking University	92.5
University of Toronto	77.3
National University of Singapore	76.4
Cornell University	80.2
University of California, Los Angeles	80.4
UCL	74.5
University of Michigan-Ann Arbor	79.3
New York University	74.2
Duke University	78.1
Northwestern University	71.7
University of Washington	71.6
Carnegie Mellon University	65.4
University of Edinburgh	66.9
Technical University of Munich	69.8

only showing top 30 rows

Рисунок 4 – Работа программы (запрос 3)

name	rank
Tsinghua University	16
Peking University	17
Fudan University	51
Shanghai Jiao Tong University	52
Zhejiang University	67
University of Science and Technology of China	74
Nanjing University	95
Southern University of Science and Technology (SUSTech)	166
Wuhan University	173
Huazhong University of Science and Technology	176
Sichuan University	196
Beijing Normal University	251-300
Sun Yat-sen University	251-300
Tongji University	251-300
Nankai University	301-350
Central South University	351-400
East China Normal University	351-400
Harbin Institute of Technology	351-400
University of Electronic Science and Technology of China	401-500
Hunan University	401-500
University of International Business and Economics	401-500
Shenzhen University	401-500
South China University of Technology	401-500
Southeast University	401-500
Southern Medical University	401-500
Tianjin University	401-500
Xiamen University	401-500
Beihang University	501-600
Beijing Institute of Technology	501-600
Changsha University of Science and Technology	501-600

only showing top 30 rows

Рисунок 5 – Работа программы (запрос 4)

name	aliases
Columbia University	Columbia University
ETH Zurich	ETH Zurich
The University of Chicago	The University of Chicago
University of Pennsylvania	University of Pennsylvania
Johns Hopkins University	Johns Hopkins University
Tsinghua University	Tsinghua University Qinghua University Tsing hua University Qing hua University
Peking University	Peking University
University of Toronto	University of Toronto
National University of Singapore	National University of Singapore nus
Cornell University	Cornell University
University of California, Los Angeles	University of California, Los Angeles
UCL	UCL ucl UCL
University of Michigan-Ann Arbor	University of Michigan-Ann Arbor
New York University	New York University
Duke University	Duke University
Northwestern University	Northwestern University
University of Washington	University of Washington
Carnegie Mellon University	Carnegie Mellon University
University of Edinburgh	University of Edinburgh
Technical University of Munich	Technical University of Munich
University of Hong Kong	University of Hong Kong
University of California, San Diego	University of California, San Diego
LMU Munich	LMU Munich Ludwig-Maximilians-Universität München
University of Melbourne	University of Melbourne UniMelb Australia
King's College London	King's College London
Nanyang Technological University, Singapore	Nanyang Technological University, Singapore NTU Nan yang
London School of Economics and Political Science	London School of Economics and Political Science
Georgia Institute of Technology	Georgia Institute of Technology
The University of Tokyo	The University of Tokyo
University of British Columbia	University of British Columbia

only showing top 30 rows

Рисунок 6 – Работа программы (запрос 5)

```

+-----+
|count(1)|
+-----+
|   1699|
+-----+

```

Рисунок 7 – Работа программы (запрос 6)

```

+-----+
|name|
+-----+
|University of Oxford|
|University of Cambridge|
|University of California, Berkeley|
|Yale University|
|University of Pennsylvania|
|Tsinghua University|
|University of California, Los Angeles|
|University of Michigan-Ann Arbor|
|Duke University|
|University of Washington|
|University of Hong Kong|
|University of Melbourne|
|London School of Economics and Political Science|
|University of British Columbia|
|KU Leuven|
|Universität Heidelberg|
|Chinese University of Hong Kong|
|Paris Sciences et Lettres - PSL Research University Paris|
|University of Illinois at Urbana-Champaign|
|University of Texas at Austin|
|Fudan University|
|Shanghai Jiao Tong University|
|The University of Queensland|
|University of Manchester|
|Wageningen University & Research|
|University of Amsterdam|
|Brown University|
|University of California, Santa Barbara|
|Utrecht University|
|Zhejiang University|
+-----+

```

Рисунок 8 – Работа программы (запрос 7)

name	subjects_offered
The University of Tokyo	Economics & Econometrics,Geology, Environmental, Earth & Marine Sciences,Accounting & Finance,Chemical Engineering,Computer Science,Communication & Media Stud
Kyoto University	Biological Sciences,Art, Performing Arts & Design,Electrical & Electronic Engineering,Communication & Media Studies,Medicine & Dentistry,Archaeology,Chemistry
Tohoku University	Law,Economics & Econometrics,Mechanical & Aerospace Engineering,Veterinary Science,Business & Management,General Engineering,Architecture,Agriculture & Forest
Osaka University	Medicine & Dentistry,Civil Engineering,Sociology,Art, Performing Arts & Design,Education,Chemistry,Languages, Literature & Linguistics,Sport Science,Other Hea
Nagoya University	Other Health,Mathematics & Statistics,Archaeology,Computer Science,Mechanical & Aerospace Engineering,Politics & International Studies (incl Development Studi
Tokyo Institute of Technology	Civil Engineering,Sport Science,Geology, Environmental, Earth & Marine Sciences,Architecture,Biological Sciences,Politics & International Studies (incl Develo
Hokkaido University	Accounting & Finance,Electrical & Electronic Engineering,Geology, Environmental, Earth & Marine Sciences,Education,Veterinary Science,Art, Performing Arts & D
Kyushu University	Languages, Literature & Linguistics,General Engineering,Economics & Econometrics,Law,Sociology,Architecture,Medicine & Dentistry,Veterinary Science,Communicat
Tokyo Medical and Dental University (TMDU)	Psychology,Mathematics & Statistics,Geography,General Engineering,Economics & Econometrics,Art, Performing Arts & Design,Other Health,Politics & International
University of Tsukuba	Law,Politics & International Studies (incl Development Studies),Business & Management,General Engineering,Archaeology,Other Health,Agriculture & Forestry,Arch
Kansai Medical University	Medicine & Dentistry,Other Health
University of Occupational and Environmental Health, Japan	Medicine & Dentistry,Other Health
Yokohama City University	Chemistry,Biological Sciences,Accounting & Finance,Medicine & Dentistry,Business & Management,Mathematics & Statistics,Other Health,Sociology,Economics & Econ
The University of Aizu	Computer Science
Fujita Health University	Biological Sciences,Medicine & Dentistry,Other Health
Hiroshima University	Archaeology,Medicine & Dentistry,Other Health,Agriculture & Forestry,Mathematics & Statistics,Architecture,Civil Engineering,Geology, Environmental, Earth & M
Juntendo University	Communication & Media Studies,Other Health,Medicine & Dentistry,Sport Science
Keio University	Art, Performing Arts & Design,Chemical Engineering,Chemistry,Communication & Media Studies,Mechanical & Aerospace Engineering,Biological Sciences,Geology, Env
Kinki University	Chemistry,Accounting & Finance,History, Philosophy & Theology,Biological Sciences,Civil Engineering,Communication & Media Studies,Other Health,Medicine & Dent
Kobe University	Education,Mechanical & Aerospace Engineering,Civil Engineering,History, Philosophy & Theology,Chemistry,Medicine & Dentistry,Physics & Astronomy,Law,Business
Kurume University	Other Health,Medicine & Dentistry,Law,Accounting & Finance,Languages, Literature & Linguistics,Business & Management,Psychology
Nippon Medical School	Medicine & Dentistry
Chiba University	Architecture,General Engineering,Sociology,Economics & Econometrics,Physics & Astronomy,Medicine & Dentistry,Biological Sciences,Psychology,Archaeology,Accoun
Hamamatsu University School of Medicine	Medicine & Dentistry,Other Health
Hyogo College of Medicine	Medicine & Dentistry,Other Health
The Jikei University School of Medicine	Other Health,Biological Sciences,Medicine & Dentistry
Kanazawa University	General Engineering,Computer Science,Sociology,Languages, Literature & Linguistics,Economics & Econometrics,Mechanical & Aerospace Engineering,Electrical & El
Kumamoto University	Geography,Other Health,Mathematics & Statistics,Electrical & Electronic Engineering,Architecture,Languages, Literature & Linguistics,Art, Performing Arts & De
Kyoto Prefectural University of Medicine	Medicine & Dentistry
Niigata University	Medicine & Dentistry,Politics & International Studies (incl Development Studies),Sociology,Other Health,General Engineering,Languages, Literature & Linguistic

Рисунок 9 – Работа программы (запрос 8)

rank_order	rank	name	scores_overall	scores_overall_rank	scores_teaching	scores_teaching_rank	scores_research	scores_research_rank	scores_citations	scores_citations_rank	scores_publications	scores_publications_rank
10	11	University of Oxford	96.4	10	92.3	15	99.7	11	99.0	28	74	174
100	10	Imperial College London	90.4	100	82.8	116	90.8	17	98.3	40	55	155
1000	99	RWTH Aachen University	63.4	1000	56.8	173	64.5	164	66.1	555	92	192
10000	801-1000	Zewail City of Science and Technology	29.8-33.9	10000	23.4	1846	16.4	932	56.6	729	41	141
1000001	Reporter	University Abdelhamid Ibn Badis Mostaganem	n/a	1000001	n/a	0	n/a	0	n/a	0	n/a	1

Рисунок 10 – Работа программы (запрос 9)

name	rank
Zagazig University	801-1000
Yeungnam University	801-1000
Yangzhou University	801-1000
University of Wyoming	801-1000
Worcester Polytechnic Institute	801-1000
University of Wisconsin-Milwaukee	801-1000
University of Winchester	801-1000
Wenzhou Medical University	801-1000
University of Warsaw	801-1000
VIT University	801-1000

Рисунок 11 – Работа программы (запрос 10)

Вывод

В ходе выполнения лабораторной работы были получены навыки работы со Spark на языке программирования Java. Был выбран датасет и сделаны 10 запросов