UDC 37:004

L. F. Panchenko, SE "Lugansk Taras Shevchenko National University"

NEW TRENDS IN DATA ANALYSIS

Panchenko L. F.

New Trends in Data Analysis

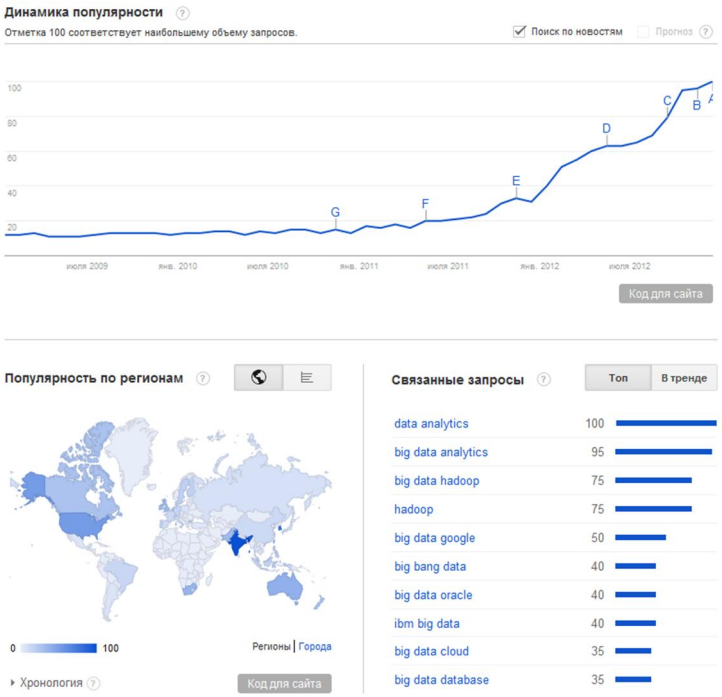
The article is devoted to the use of cloud technology of data analysis. The opportunities of programs for the academic community SAS On Demand for Academics are discussed. This program allows to organize the work of a teacher and a student in the course of data analysis with cloud computing: the client part works on the computers of teachers and students, and the analysis is carried out with cloud servers SAS. The advantages of this approach for university courses are analyzed: free, using of advance data mining techniques, the skills formation to work with cloud technologies, the ability to have a cloud data sets for joint analysis.

The content and organization of learning in the course of data analysis "Passion Driven Statistics" of project coursera.org are considered. Course syllabus consists: statistical software set up; data sets and data documentation; data management; descriptive statistics and data visualization; comparing means; tests of categorical independence and correlation. The different data sets (characteristics of Mars craters, adolescent health, adult psychiatric, social, economic and health indicators of countries worldwide) to analyze are proposed. Evaluations included quizzes (40%), applied data assignments (20%) and a final project via individual student blog sites (40%). Learning materials has been presented as text, video lectures and demonstrations.

Analysis of the clouds of words with student's answers shows that students relate concepts of data analysis with "learning", "new skills formation", "understanding", "improvement", "refreshing".

Key words: data analysis, cloud computing, education, SAS.

Among the most significant trends of 2013 by research agency Gartner [6] – are processing large data sets ("big data") and cloud calculation. Pic. 1 shows the dynamics of the search query "big data", obtained with the help of service Google Ngram. The requests such as "analysis", "big data" are described as excess popular. Therefore, an important task of High School is to prepare a competitive specialist knowledgeable in the corresponding technologies.



Pic.1. Dynamics of popularity of search request big data

Educators, followed by representatives of business organizations, are interested in using cloud services. Cloud service Office 365 from Microsoft provides a platform for information systems “open education”, which is aimed at automating the process of education in Russian schools. With the help of such services, the educational institutions will be able to arrange a full distant training, operative communication with the parents, electronic administrative accounting [7]. Recently, the All-Ukrainian scientific-methodical seminar was dedicated to the use of cloud technologies. [4]. Seminar highlighted the issues in cloud technologies, their use in outdoor education, high school, IHE, postgraduate education, also there were examples of cloud-based ways of learning physics, mathematics and computer science. The analysis of publications of reports [4] and personal experience of seminar [2], are helping to make a conclusion about the lack of developing issues on the use of cloud technology in relation to the subject area “data analysis”.

Purpose of the article: to consider the possibilities of cloud technology in relation to the field of data analysis.

Analysis of the usage of statistical software in the master`s programs of Western universities, particularly in the social and humanitarian sciences [5], shows

the following dynamics using the most popular packages (in% of total number of courses):

Table 1

Dynamics of usage of statistical software in the master's programs of universities [5]

Software	1996-2000	2001-2005	2006-2010
SPSS	33	31	31
Excel	30	29	17
SAS	33	32	24
STATA	4	8	27

At the same time, employers require from their employees the ability to use: just SPSS – 20 %, SPSS or SAS – 22%, only SAS – 19 %, SAS or STATA – 9 % [5]. Thus, SAS and SPSS are the most popular statistical field. The course, which was developed by the author and O. Adamenko with the computer data analysis for training various specialties, combines the tasks from data analysis in fields of Excel and SPSS. While finding the expanse ways to use the software on the statistical analysis of the data, we turned to the academic community support programs of SAS. This software (SAS OnDemand for Academics) includes a free opportunity for students in courses while using SAS and academics who teach courses with the support of SAS, download to your PC client and SAS to analyze data using cloud technologies using Server SAS [10].

The sequence of steps to use this opportunity for teachers includes registration and creating of the private profile SAS, election of required client`s software, downloading it to your computer, registration the course, sending out information to students about the future course. After all these steps, we got a free certificate for e-learning opportunities of SAS for 1 year. E-learning courses include: "Requests and Reports", "Fundamentals of Programming", "Programming: Data Management technique", "analysis of variance, regression, logistic regression" of support among

the SAS Enterprise Guide; and “Applied Analytics” and “predictive forms for business intelligence” in SAS Enterprise Miner.

“Passion Driven Statistics” course which was held on Coursera in March 2013 was also dedicated to Data analysis with the use of SAS [9]. As a basic statistical software it was also used SAS OnDemand for Academics. The Sillabus of the course included the following topics:

- Installation of statistical software, data sets and its documentation.
- Data management. Descriptive statistics and data visualization.
- Comparison of average (ANOVA), tests of independence (Chi Square) and correlation.
- Moderation.
- Presentation of statistical results.
- Students activity in this course includes:

Research documentation of data;

– fulfilment of the basic programs;

– data management;

– charting and diagramming;

– hypothesis testing;

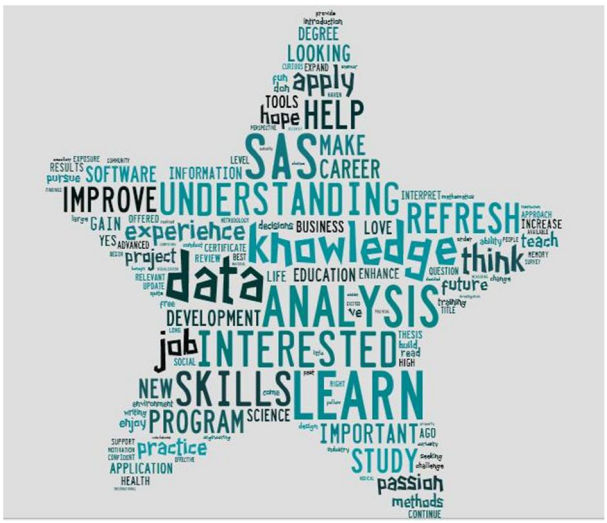
– result of the study

Students were given of several sets of data from different areas: 1) the characteristics of the craters of Mars, 2) teenagers health, and 3) adult mental health and its disorders, and its frustration 4) social, economic and health numbers of the world.

Interestingly, all the data were originally presented in the format of SAS. And then the discussion on the forum was also extended their analysis in SPSS, R, Stata, and presented in a csv-file. Based on the student’s choice of data, everyone had generated statistical hypothesis for testing, prepared data for analysis, conducted descriptive and analytical analysis, evaluated, interpreted and presented research results.

Assessment of students in the course is consisted of tests, tasks of data analysis and presentation of the final results by the individual blogs of students. During tests the student got 40%, for the task of data analysis – 20%, final project estimated at 40%.

Supporting materials were presented in the form of text also video lectures and demonstrations. List of the best blogging was introduced to other students for acquaintance [8].



Pic. 2. Cloud of words drawn from the responses of students [8]

Analysis of word clouds from Pic. 2 shows that students associate the concept of data analysis from the “studying” formation of “essential skills”, “understanding”, “improvement”, “restoration”. In the Lisa Dirker`s blog we found a wide list of reasons why students join the course [8]. Here are a few:

I tried to find myself! I love statistics.

I believe that an understanding of statistics provides a solid foundation for high citizenship in the era of “big data”.

I like the format, namely the applying of knowledge for solving the problems.I like to tell stories through statistics. To build a “Start Up”.

Get the required skills, to become a scientist in the field of data analysis.

Suffer from boredom of the program of higher school. It’s free!

In order to study first-hand experience of pedagogy it`s used in MOOC by teacher.

I am a teacher of applied statistics, and I want to know more about the statistics and find interesting ideas for their use in the classroom.

Understanding the data in new ways will serve me not only professionally, but also creatively.

I love data!

Thus, cloud technologies of data analysis are combining the leading trend development of information technology: analysis of large data sets, so-called “big data” and cloud calculating. Resources programs which are support the academic community SAS OnDemand for Academics, allow organizing the work of students and teachers in the course of data analysis using cloud technologies. The client`s side runs on computers of teachers and students, and the analysis is carried out using cloud servers SAS. The benefits of this approach for university courses: free of charge, the use of advanced data analysis, formation of skills to work with cloud technologies, the ability to place the data sets for joint analysis in “clouds”.

Further areas of research are: the expansion of the program and teaching of courses on data analysis for university students through the use of SAS OnDemand for Academics and cloud technologies.

References

1. **Adamenko O. V.** Teoritichno-metodichny zasady navchannya studentiv analizu z vikorystannyam komp`utera [Theoretical and methodological basis for the analysis of student learning data using computer]: Bulletin of. Lugansk Taras Shevchenko.National University – № 17. – 2010. – P. 31 – 35.

2. **Adamenko O. V.** Hmarni tehnologiy analizu dannih [Cloud technologies of data analysis]: Hmarni tehnologiy v osviti: Mater.Vseukr. nauk.-metod.sem. Kriviy Rig: Vidavnichiy viddil. – 2012. – P. 143 – 144.

3. **Panchenko L. F.** Komputerniy analiz dannyh: ucheb. Posobiye dlya studentov vish. ucheb. zavedeniy [Computer Data Analysis: tutorial for students from

higher institutions]: SE “Lugansk Taras Shevchenko.National University”. – Lugansk : “SE LNU named after Taras Shevchenko”. – 2010. – 188 p.

4. **Hmarni** tehnologiy v osviti [Cloud technologies in education]: Mater. Vseukr. Nauk.-metod. Sem. Kriviy Rig: Vidavnichiy viddil KMI. – 2012. – 173 p.

5. **Adams W. C.** Statistical Software for Students: Academic Practices and Employer Expectation /William C. Adams, Donna Lind Infeld, Carli M. Wulff. – [Electronic resource]. – Access mode : http:// umdcipe.org/conferences/Classroom/Presentations/APPAM%20Teaching%20Workhop_Adams%20Infeld_pre s_11_1109.pdf

6. **Gartner:** Top 10 Strategic Technology Trends For 2013 [Electronic resource]. – Access mode : http://www.forbes.com/ sites/ericsavitz/2012/10/23/ gartner-top-10-strategic-technology-trends-for-2013/

7. **Microsoft** Office 365 [Electronic resource]. – Access mode : //www.ferra.ru/ru/techlife/news/2013/05/27/Microsoft-Office-365-ru-school/

8. **Passion** Driven Statistics. – [Electronic resource]. – Access mode : http://passiondrivenstatistics.tumblr.com/

9. **Passion** Driven Statistics. – [Electronic resource]. – Access mode : https://www.coursera.org/course/pdstatistics

10.**SAS** OnDemand for Academics [Electronic resource]. –Accessmode : https://support.sas.com

Панченко Л. Ф.

Нові тренди аналізу даних

Стаття присвячена питанням використання хмарних технологій аналізу даних. Виявлено можливості програми підтримки академічної спільноти SAS OnDemand for Academics, які дозволяють організувати роботу викладача йстудента в курсах аналізу даних із використанням хмарних технологій: клієнтська частина працює на комп’ютерах викладачів та студентів, а сам аналіз здійснюється за допомогою хмарних серверів SAS. Аналізуються переваги такого підходу для університетських курсів: безкоштовність, використання передових технологій аналізу даних, формування умінь і навичок роботи з хмарними технологіями, можливість розташовувати у хмарах масиви даних для спільного аналізу . Розглядається зміст та організація навчання в курсі з аналізу даних „Passion Driven Statistics” проекту coursera.org.

Ключові слова: аналіз даних, хмарні технології, освіта, SAS.

Панченко Л. Ф.
Новые тренды в анализе данных
Статья посвящена вопросам использования облачных технологий анализа данных. Определены возможности программы поддержки академического сообщества SAS OnDemand for Academics, которые позволяют организовать работу преподавателя и студента в курсах анализа данных с использованием облачных технологий: клиентская часть работает на компьютерах преподавателей и студентов, а сам анализ осуществляется с помощью облачных серверов SAS. Анализируются преимущества такого подхода для университетских курсов: бесплатность, использование передовых технологий анализа данных, формирование умений и навыков работы с облачными технологиями, возможность располагать в облаках массивы данных для совместного анализа. Рассматривается содержание и организация обучения в курсе анализа данных „Passion Driven Statistics" проекта coursera.org .
Ключевые слова: анализ данных, облачные технологии, образование, SAS.

Information about the author

Lubov Feliksovna Panchenko – PhD, Professor of Theoretical and Applied Informatics Department of Luhansk Taras Shevchenko National University. Among scientific interests are: information and communication technologies in education, information and educational surroundings of the University, the statistical data analysis.

The article was received by the Editorial Office on 08.04.2013
The article was put into print on 26.04.2013