DOI: https://doi.org/10.53052/9788367652018.02

Aleksandra KŁOS-WITKOWSKA[1], Marcin BERNAŚ[1], Krzysztof WITOS[1], Eugenia KOVATCHEVA[2], Vasyl MARTSENYUK[1], Georgi DIMITROV[2], Dejan RANCIC[3], Oleksii BYCHKOV[4]

BADANIE ISTNIEJĄCYCH PROGRAMÓW NAUCZANIA W ZAKRESIE BIG DATA

Streszczenie: Przegląd został przygotowany na podstawie informacji reklamowych o szkoleniach z zakresu technologii Big Data i Data Science. Badanie zostało przeprowadzone on-line za pomocą narzędzi formularzy Google. Ze względu na różne formaty i specyfikację informacji reklamowych w każdym przypadku dane zostały zebrane przez naukowców na podstawie wyszukiwania fraz. Wnioski pokazują nam podstawowe umiejętności twarde i miękkie oraz inne wymagania, które należy wziąć pod uwagę przy projektowaniu szkolenia z Big Data

Słowa kluczowe: iBIGworld, Big Data, szkolenie, umiejętności twarde, umiejętności miękkie

RESEARCH OF EXISTING TRAINING PROGRAMS IN THE FIELD OF BIG DATA

Summary: The survey was prepared based on the advertising information about training courses in the field of Big Data technology and Data Science. The survey was performed on-line using google forms tools. Due to various formats and specification of advertising information in each case the data was collected by scientists based on phrase search. Conclusions shows us the basic hard and soft skills and other requirements that should be taken into account when designing the training course on Big Data

Keywords: iBIGworld, Big Data, training course, hard skills, soft skills

1. Introduction

Last time the development of a training course on Big Data is the subject of consideration in some works [1-4]. In turn, when designing the course, we should use

[1] Department of Computer Science and Automatics, University of Bielsko-Biala, Poland: (awitkowska, mbernas, kwitos, vmartsenyuk)@ath.bielsko.pl
[2] University of Library Studies and Information Technologies, Sofia, Bulgaria: g.dimitrov@unibit.bg, e.kovatcheva@unibit.bg
[3] University of Niš, Nis, Serbia: dejan.rancic@elfak.ni.ac.rs
[4] Taras Shevchenko National University of Kyiv, Kiev, Ukraine: oleksiibychkov@knu.ua

the experience of the existing ones [5-16]. Hence in the given work, we would try to collect the good practices of the training courses in the field of Big Data. This survey was carried out in the context of project no. 2020-1-PL01-KA203-082197 entitled "Innovations for Big Data in a Real World". The survey was obtained by the scientists based on the advertising information about training courses in the field of Big Data technology and Data Science. The survey was performed on-line using google forms tools. Due to various formats and specification of advertising information in each case the data was collected by scientists based on phrase search. The phrases used were "Training AND Course AND Big Data AND Data Science AND Analytics AND Analysis AND Processing AND Machine Learning" statement. The survey was performed during a period from the 1'th of September 2020 to 28'th of February 2021. To obtain wide range of data, multiple choice question fields, with additional open question field were offered to overcome the shortcomings of narrowed answers suggestions. The survey contains both open and closed questions. To make a process of data collection un-biased no additional recommendation was added. No events were reported during that time that could influence the result. The training courses were selected randomly from the advertising sites. The web sites were selected based on popularity and number of references. The survey data were presented in a quantitative form. The data was aggregated to the category based on syntactic analysis I.e., the difference in form and not meaning .The open descriptions were analysed with the help of bar charts of word frequencies, word clouds, and word associates .

2. Collection and analysis of data

The research was analysed on the basis of 65 questionnaires carried out in 4 countries participating in the project: Poland , Ukraine, Bulgaria Serbia.
The study described research of existing training programs in the field of Big. Analysing below diagram ,we can see that the highest number of questionnaires came from Serbia-23 (35.4%) and Bulgaria-22(33.8%), while 13(20%) of questionnaires came from Poland and 7(10.8%) from Ukraine (Fig. 1).

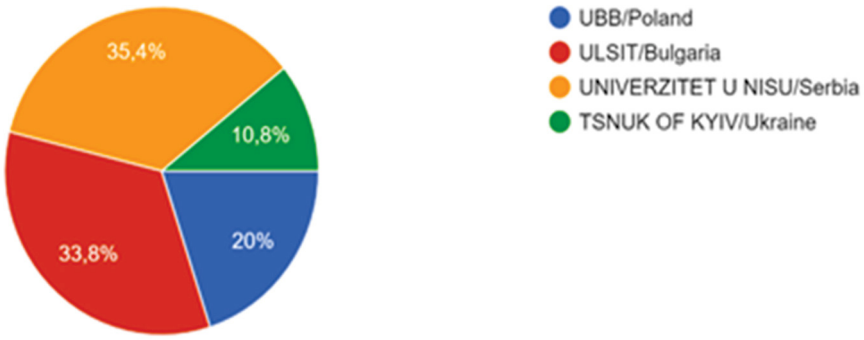

Figure 1. Distribution of questionnaries among the project partners

3. Results

3.1. Country and Organization in which the training takes place

The next question was about Country and Organization in which the training takes place. The goal was to analyze the countries which offer the most of courses in the field of Big Data and in tourn have the most experience.

Data description.
The survey showed that most courses are organized in the U.S. 16.9% and also in Serbia 12,3% , United Kingdom 10.8% and Ukraine 10.8%. Also courses are organized in Germany, Poland, Finland, Greece, Spain, France, Switzerland, France, Italy, Netherlands, Canada, Bulgaria, Serbia, Ukraine. The percentage share is presented in the Table 1 and Fig. 2.

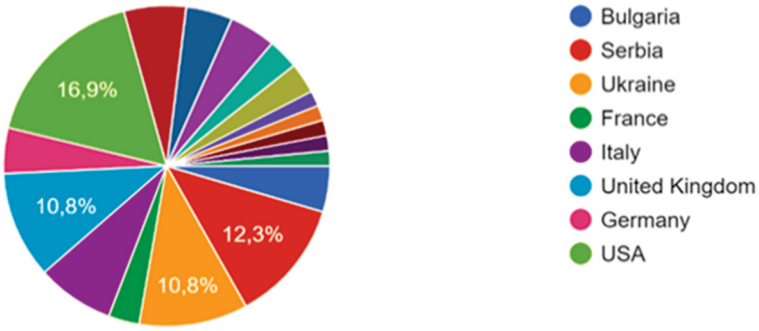

Figure 2. Distribution of the investigated courses among the countries

Table 1. Country in which the training takes place

Country	percent	number of responses
United Kingdom	10,8%	7
Germany	4,6%	3
USA	16,9%	11
Poland	6,2%	4
Finland	4,6%	3
Greece	4,6%	3
Spain	3,1%	2
Switzerland	3,1%	2
Other countries	1,5%	1
International organization	1,5%	1
Belgium,Germany, Spain, Netherlands, France, Italy	1,5%	1
Netherlands	1,5%	1
Canada	1,5%	1
Bulgaria	4,6%	3
Serbia	12,3%	8
Ukraina	10,8%	7
France	3,1%	2
Italy	7,7%	5

Discussion

For the reasons given we conclude that the most of courses in the field of Big Data are organized in USA, UK and EU.
Main conclusions:

- The survey shows that the training courses in the field of Big Data organized in USA, UK and EU are promoted and accessible for clients for the best.
- We need to focus on the experience in organizing and conducting in training courses in the field of Big Data obtain from USA, UK, EU and other countries (Serbia and Ukraine).

3.2. Type of the course

The aim of the question was to study the types of training courses offered in the field of Big Data

Data description

Analyzing the types of courses (Fig. 3), it was found that the most courses are organized at the level of master degree -42(64.6%),followed by bachelor degree - 11 (16.9%), course of academy – 6 (9.2%), training course - 5 (7.7%) and short course – 1 (1.5%).

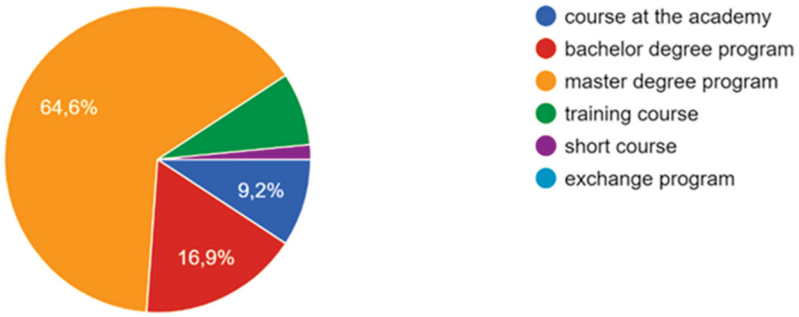

Figure 3. Distributions of the types of the Big Data courses

Discussion

The results of survey showed the distribution among the types of training courses offered, displaying the dominance of the courses during master study, whereas the short courses are offered rarely.

Main conclusions:

Training courses in the field of Big Data study require primary study in the field of IT. So, the most of training courses are offered within master degree programs.
Short courses are less preferred as the time requirements are not satisfied to present wide range of topics needed for Big Data.

Bachelor and Master programmes in the field of Big Data are preferred in comparison with courses at the academy, since they are advanced continuation of traditional IT courses.

3.3. Level of the course

The aim of the question was to study the levels of the courses offered in the field of Big Data.

Data description

Looking at the results of the analysis of course levels (Fig. 4), we can say that most courses are organized at advanced level - 37 (56.9%) and at beginners level – 26 (40%). The bottom of the list is intermediate level – 1 (1.5%) and master level - 1 (1.5%).

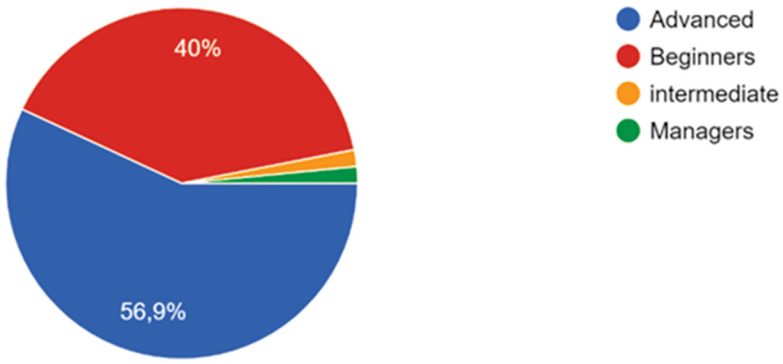

Figure 4. Distributions of the course levels

Discussion

The survey shows the dominance of advanced and beginning level courses which are the most needed.

Main conclusions:

- Big Data technology assumes the knowledge and skills of the basic IT technologies. So, the most of propositions of training courses are offered at the advanced level.
- Big Data is one of novel technology. Therefore, a lot of persons are novices in this branch, requiring the courses at the level for beginners.
- The are definitely less proposition of training courses at the level of intermediate and managers, which shows the great opportunity and perspective when developing such level courses.

3.4. Name of the training course/discipline

The aim was to discover main tendencies when naming the training courses in Big Data.

Data description

When the respondents were asked to fill in the name of the training course/discipline, they indicate the names of 65 courses. When processing these data we have primarily transformed the text (cleaning the text data making transformations like removing special characters from the text, removing the unnecessary whitespace and converting the text to lower case, removing the stopwords, text stemming by reducing the word to its root form).

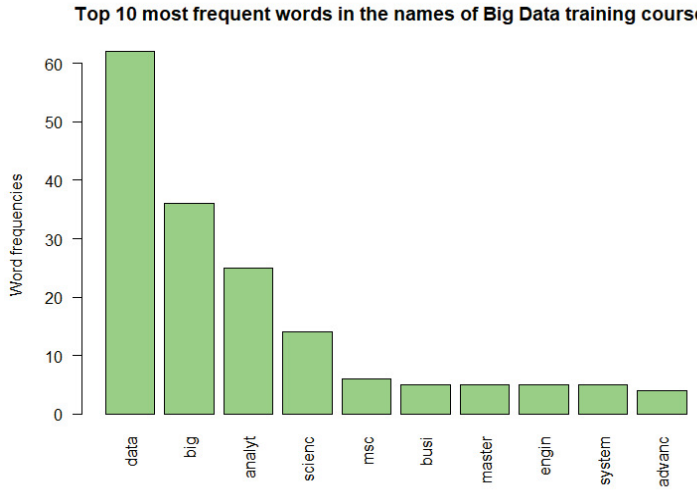

Figure 5. Bar chart of the top 10 most frequent words within the names of the training course/discipline

One could interpret the following from the bar chart (Fig. 5):

- the most frequently occurring word is "data",
- "big", "analytics" and "science" are the next four most frequently occurring words, which indicate that most often they pay attention on analytical aspects and data scienc within the courses dealing with Big Data.

Finally, the word "data" for notions like "data set", "data science", "data scientist", etc. is also on the chart, and you need further analysis to infer in what context is used A word cloud in Fig. 6 helps us to visualize and analyze qualitative content within names of training courses. The image is composed of keywords found within the names, where the size of each word indicates its frequency in the titles.

Figure 6. Word cloud plot within the names of training courses

The word cloud shows additional words that occur frequently and could be of interest for further analysis. Words like “system”, “busi” (root for “business”), “engin” (root for “engineer”), etc. could provide more context around the most frequently occurring words and help to gain a better understanding of the main titles.

Correlation has demonstrated whether, and how strongly, pairs of words are related. This technique was used effectively to analyze which words occur most often in association with the most frequently occurring words in the names of training courses, which helps to see the context around these words.

We can see the following results of word association..

The output indicates that “intelligen” (which is the root for words “intelligent”, “intelligence”) and “architectur” (which is the root for word “architecture”) occur 43% of the time with the word “system”. You can interpret this as the context around the word (“system”) is dealing with artificial intelligence or system architecture. Similarly, the root of the word “engineering” is highly correlated with the word “software”. This indicates that most responses are saying that training courses in Big Data in branch of engineering are primarily dealt with software engineering context.

Finding terms associated with words that occur at least 10 times or more, we get the comprehensive results on the content of the courses.

Discussion

The analysis of the survey question allowed us to discover main tendencies in the titles of training courses in Big Data.

Main conclusions:

- the training course should include the term “Big Data” clearly;
- it is desirable to show in the name of the course the connection with data science and analytics;
- some in formation on the level of the program (e.g. bachellor, master or at least advanced) should be indicated in the name of the training course.

3.5. Entry-level for the course - requirements for enrolling

The aim of this question was to study main requirements for enrolling for the training course in Big Data

Data description

When the respondents were asked to fill in the prerequisites of the training course/discipline, they indicate the requirements for enrolling of 65 courses. When processing these data we have primarily transformed the text (cleaning the text data making transformations like removing special characters from the text, removing the unnecessary whitespace and converting the text to lower case, removing the stopwords, text stemming by reducing the word to its root form).

One could interpret the following from the bar chart on Fig. 7:
- The most frequently occurring words are “basic” and “knowledge”.
- “program”, “python”, “algorithm”, “rlanguage” and “statist” are the next five most frequently occurring words, which indicate that most often they pay attention on requiring practical skills in programming and statistics obrained prior the courses dealing with Big Data.

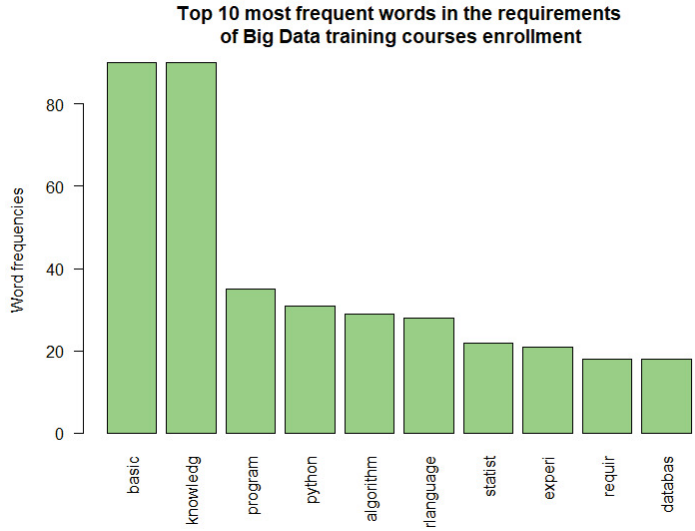

Figure 7. Bar chart of the top 10 most frequent words within the prerequisites of the training course/discipline enrollements

Finally, the words “database” and “data” for notions like “big data”, “data set”, “data science”, etc. are also on the chart, and you need further analysis to infer in what requirement are needed.
A word cloud in Fig. 8 helps us to visualize and analyze qualitative content within requirements of training courses enrollment. The image is composed of keywords found within the prerequisites, where the size of each word indicates its frequency in the term.

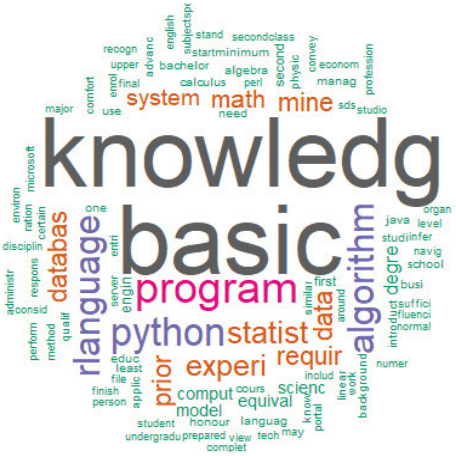

Figure 8. Word cloud plot within the prerequisites of training courses

The word cloud shows additional words that occur frequently and could be of interest for further analysis. Words like "system", "math" (root for "mathematics"), "mine" (root for "mining"), etc. could provide more context around the most frequently occurring words and help to gain a better understanding of the main prerequisites.
Correlation has demonstrated whether, and how strongly, pairs of terms are related. This technique was used effectively to analyze which words occur most often in association with the most frequently occurring words in the prerequisites of training courses enrollment, which helps to see the context around these words.
We can see the results of word association as shown in Tables 3 and 4.

Discussion

As a result of analysis of the answers basing on word frequencies, word cloud and associations, we have studied the main prerequisites required for enrollment for training courses.

Main conclusions:
The prerequisites have to include
- the training course in Big Data has to assume certain level of programming skills;
- requirements for enrolling should assume college level of knowledge of Python, R, SDS, etc.;
- requirements for enrolling should assume college level of calculus and algebra;
- skills of using databases (SQL and NoSQL) are to be required for enrolling for the Big Data course;
- the ability of algorithmic approach (e.g. data mining) to solve the problems dealing with the data is to be required for enrollment for Big Data course.

3.6. Topics covered in the course

The aim of this survey question is to investigate the main topics included in the Big Data training courses and to infer which ones are reasonable to cover in the planned course.

Data description

When the respondents were asked to fill in the topics covered in the training course/discipline, they indicate the topics included in 65 courses. When processing these data we have primarily transformed the text (cleaning the text data making transformations like removing special characters from the text, removing the unnecessary whitespace and converting the text to lower case, removing the stopwords, text stemming by reducing the word to its root form).

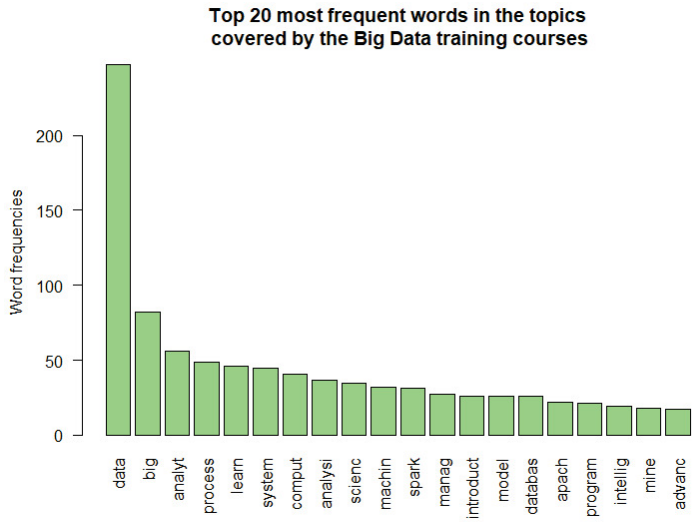

Figure 9. Bar chart of the top 20 most frequent words within the topics covered with the training course/discipline on Big Data

One could interpret the following from the bar chart on Fig. 9:

- The most frequently occurring words are "data" and "big".
- "analytics", "processing", "learning", "computing", "analysis", "managing", "modeling", "mining" are the next most frequently occurring words, which indicate that most often they pay attention on various stages of Big Data technology within the topics covered by training courses/disciplines.
- "science", "intelligence", "learning", "machine" mean the importance of the topics dealing with data science, artificial intelligence and machine learning.
- "spark", "apache" mean tools considered within the courses.

Finally, the words "program" and "advanced" are also on the chart, and you need further analysis to infer in what context they are used.

A word cloud in Fig. 10 helps us to visualize and analyze qualitative content within topics covered in the training courses. The image is composed of keywords found within the topics, where the size of each word indicates its frequency in the term.

Figure 10. Word cloud plot basing on the topics covered within the training courses

The word cloud shows additional words that occur frequently and could be of interest for further analysis. Words like “system”, “math” (root for “mathematics”), “mine” (root for “mining”), etc. could provide more context around the most frequently occurring words and help to gain a better understanding of the main prerequisites.
Correlation has demonstrated whether, and how strongly, pairs of terms are related. This technique was used effectively to analyze which words occur most often in association with the most frequently occurring words in the topics of training courses, which helps us to see the context around these words.
Rave also obtained the results of word association.

Discussion

Basing on the study of rensponses for the survey question, we have investigated the main topics of existing courses in the field of Big Data.
Main conclusions:
The planned course in the field of Big Data has to include the following topics:

- system architecture of large-scale system using Big Data;
- Big Data platforms: Apache Hadoop Ecosystem and Their Components;
- distributed streaming data-flow computing with Big Data (Apache Flink, Apache Storm);
- cloud computing (Hadoop Mapreduce and Apache Spark);
- software bus and stream processing for Big Data solutions (Apache Kafka);
- analytical processing of Big Data (Apache Spark);
- dataflow between systems; using (Apache NiFi);
- business analytics based on Big Data;
- architecture of microservices for Big Data solutions;
- algorithms of analytics and machine learning for Big Data;
- intelligent (e.g. Bayesian) analysis of large-scale data;
- deep learning for Big Data;
- forecasting and risk assessment for large-scale data;
- high performance analysis of large-scale data (e.g. world econometric, geospatial);

- visualizing Big Data.

3.7. Short topics" description could be extended by the following elements

The aim of the question was to search the most important elements that should be included in the short topics description of the training courses in the field of Big Data.

Data description. Looking of the results of the responses we see the following distribution of propositions to extend the short topics description: with " What direct Big Data topics will be presented "- 23 (35,4%),"Ways to process large volumes of data using hierarchical storage, hashing and filtering"- 13 (20%), "Ways to select the efficient algorithm to Big Data, which takes under consideration its scale"- 8 (12,3%), "Ways to select appropriate sampling and filtering method for given Big Data analysed case"- 4 (6,2%), "Ways to tackle with concurrency / parallelism problems of Big Data scale"- 12 (18,5%), "Drive better business decision with an overview of how Big Data is organized, analysed and interpreted"- 1 (1,5%)".

Complementary short courses may be completed by elements presented on Fig. 11:

Figure 11. Elements to complete short courses

Discussion

For the reasons given we see that the most important extensions of the training courses descriptions are concerning with the direct Big Data topics, the ways of processing and efficient algorithms related to Big Data.

Main conclusions:

The advertising of the training courses offered in the branch of Big Data should include some extended descriptions of the topicscovered in the course

When advertising training course, the extended descriptions of the topics covered should include both direct Big Data topics and specific topics related to the ways of processing large volume of data, namely, taken into account data structure, large scale, efficiency, parallelism, etc.

3.8. What related topics will be presented

The aim of the question was to state the related topics that have to be presented.

Data description

Related topics, presented in the table and graph below was also analyzed.
We could see the most important to responders (47 answers) was visualize results of analysis of Big Data72,3%, also skills for obtaining information from existing sources (46 answers) 70,8% and ability of using wide range of Big Data analytics platforms (45 answers) 69,2%. The rest of the results are presented in the Table 2 and Fig. 12.

Table 2. Related topics to be presented

No.	Related topics	Percentage	No of responses
1	Obtain information from existing sources (streaming data/ historical ones/ applications logs/ open-source databases	70,8%	46
2	Effectively use variety of data analytics techniques (Machine Learning, Data Mining, Prescriptive and Predictive Analytics).	56,9%	37
3	Apply quantitative techniques (statistics, time series analysis, optimization, and prediction)	53,8%	35
4	Process heterogeneous data (natural language, visual objects, data, text and other).	60,0%	39
5	Visualize results of analysis of Big Data	72,3%	47
6	Deploy solution (merging data collection, storage, analysis and visualization)	43,1%	28
7	Using wide range of Big Data analytics platforms	69,2%	45
8	Develop and operate large scale data storage (e.g., Data Lakes, Hadoop and others)	55,4%	36
9	Apply data security mechanisms and controls at each stage of the data processing	21,5%	14
10	Design, build, operate relational and nonrelational databases (SQL and NoSQL)	56,9%	37
11	Process large dataset (i.e., ETL, OLTP, OLAP)	40,0%	26
12	Ensure data quality, accessibility, interoperability, compliance to standards, and publication	30,8%	20

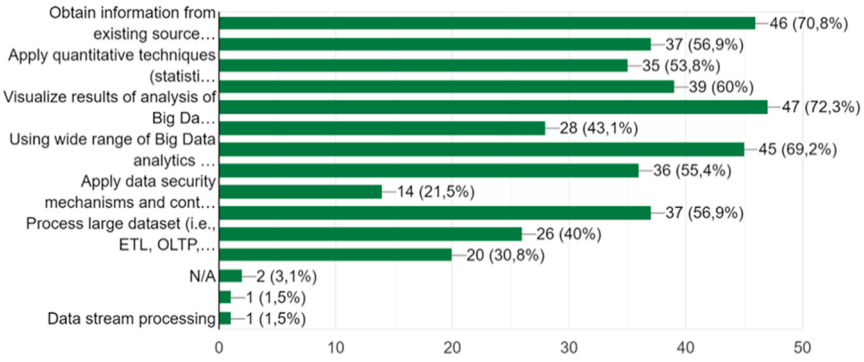

Figure 12. Related topics to be presented

Discussion

For the reasons given we can see that the most important related topics that need to be included in the training courses.

Main conclusions:

The topics of visualization and presentation of Big Data are of important.

The training courses should include good practices of processing Big Data gathered form real application.

The topics of using analytics tools should be included in the training courses in Big Data.

The developing and using data storages based on e.g., Data Lakes, Hadoop and others have to included in the training courses.

The data security was mensioned as the less important when considering the ralated topics of Big Data.

3.9. Expected results

The aim of the question was to obtain information about expected outcomes obtain after taking the course.

Data description

Based on the survey, it can be seen the greatest interest in master degree - 41 (63,1%), bachelor degree - 11 (16,9%), professionals certificate – 4 (6,2%) and postgraduate diploma - 4 (6,2%), whereas less interest was noted for training certificate - 1 (1,5%), Master's certificate - 1 (1,5%), credits for course - 1 (1,5%) (Fig. 13).

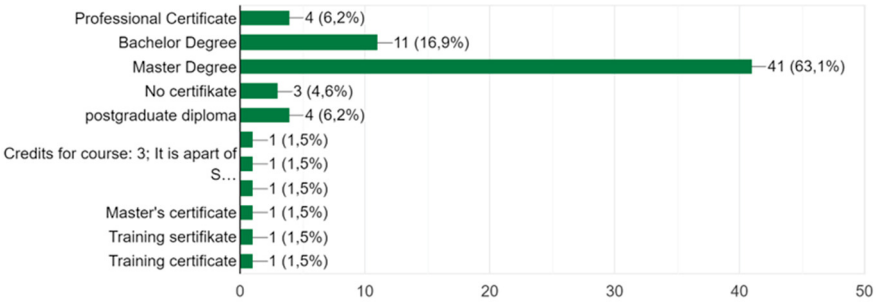

Figure 13. Expected results after the course

Discussion

For the reasons given we can see that there is a great interest in obtaining outcomes after the taking courses in the Big Data field.

Main conclusions

When planning and developing training courses in the field of Big Data when should take into account that:

- the highest interest was noticed to receive a full academic education. A master's degree was the most preferred, but a bachelor's degree was also in high demand,
- it was noted that certificate and postgraduate diploma are the less interest for professionals,
- certificates (training and masters) and credits courseswerethe least popular among existing training courses in Big Data.

3.10. Form of knowledge assessment

The aim of the question was to obtain information about form of knowledge assessment (Fig. 14).

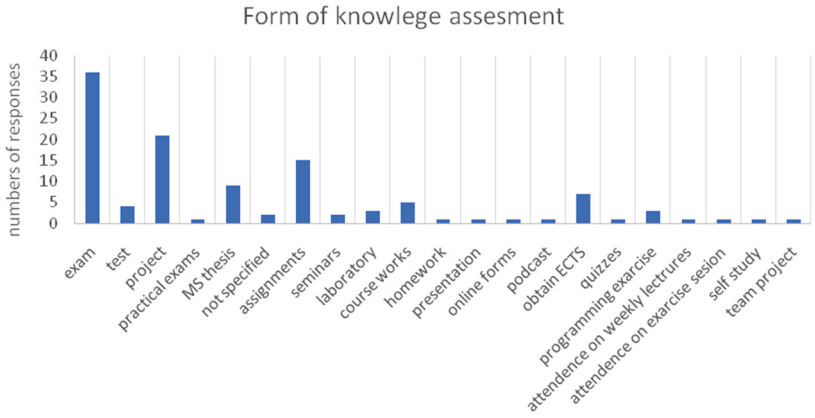

Figure 14. Forms of assessment

Data description

Form of knowledge also has been analysed. It has been obtain that the most common form of testing is an exam, next project and assignments (Fig. 15). Also obtaining the appropriate number of ECTS credits and master's thesisare important. Other forms are used less frequently (test, practical exams, seminars, homework, presentations, podcast, quizzes, programming exarcises, self-study or team projects (Fig. 16).

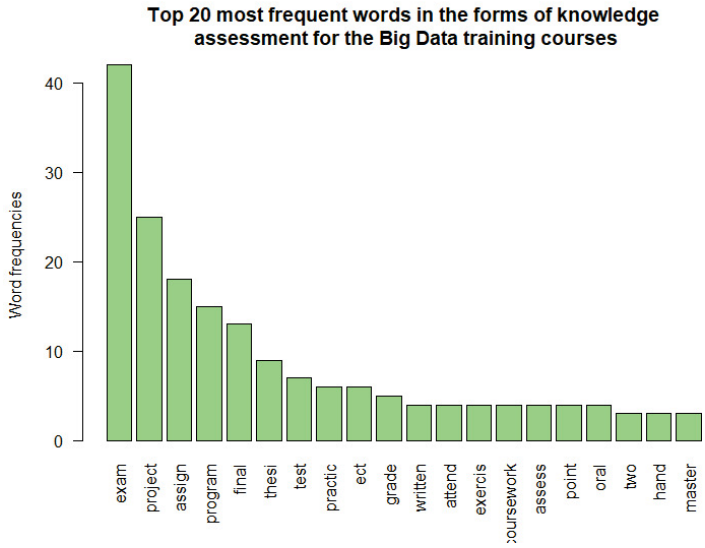

Figure 15. Analysis of the forms of assesment

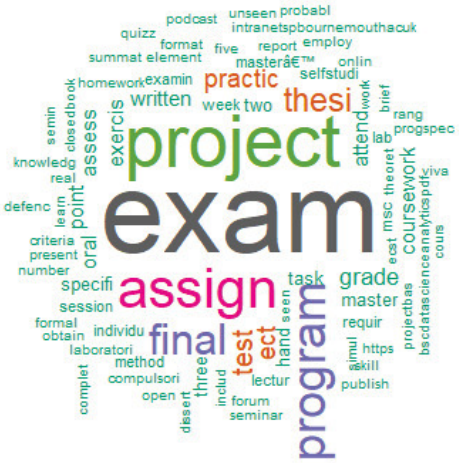

Figure 16. Word cloud plot basing

Discussion

Basing on statistical processing of (histogram, bar chart, word cloud), we can see that many different types of knowledge assessment were considered to test the learning progress and the results of knowledge gained.

The most preferred form were exams, projects, assignments and obtaining the appropriate numbers of ECTS points.

Main conclusions:
- The most preferred form of knowledge assessment are: exams, projects, assignments and obtaining the appropriate numbers of ECTS points.
- MS thesis, course work are ranked in the middle.
- The less popular form of knowledge assessment are: practical exams, seminars, laboratory, homework, presentation, online forms, podcast, quizzes, programming exercise session, self-study and team projects.

3.11. Duration of training (in days, months or years)

The aim of the question was to obtain information about duration of training.

Data description

Taking into accound duration of training it could be seen that the most courses are offered for one semester, but also popular are courses for one year and four years (Fig. 17).
It also was found traing for 1,5 year, 1 day,13 lectrures, 14 classes, 14 weeks, 2years, 220 hours, 3 days, 3 months, 3 semesters, 4 semesters, 4 years, 6/8 weeks, 8 months, 8 weeks, 9 months, 4 months, 12 months, 1,5-3 years but those were not that popular.

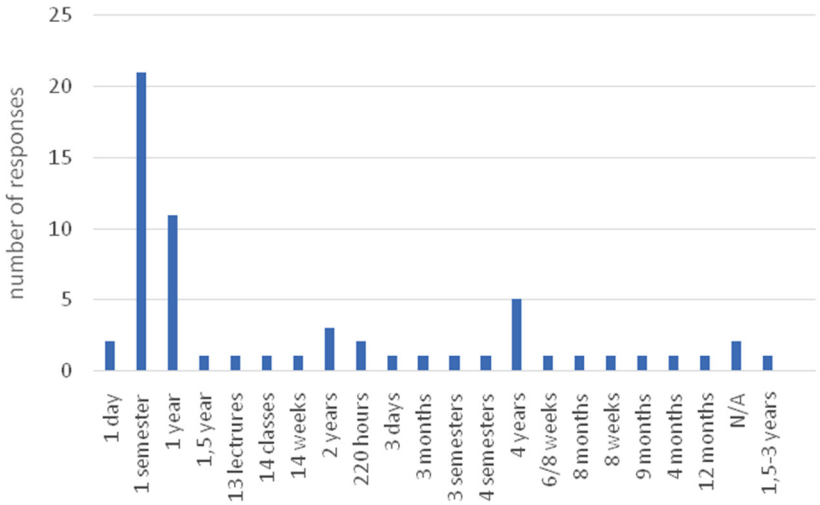

Figure 17. Duration of training

Discussion

For the reasons given we can see that there is a large variety of course durations. Available both short one-day courses as well as courses lasting 4 years. However, the most of courses are for 1 semester.

Main conclusions:
When planning and developing training courses in Big Data, one should bear in the mind the following aspects of the duration of course:

- the most of courses are offered for one semester, but also popular are courses for one year and four years,
- trainings for 1,5 year, 1 day,13 lectrures, 14 classes, 14 weeks, 2years, 220 hours, 3 days, 3 months, 3 semesters, 4 semesters, 4 years, 6/8 weeks, 8 months, 8 weeks, 9 months, 4 months, 12 months, 1,5-3 years are not so popular.

3.12. Price (in euro)

The aim of the question was to obtain information about price of the courses in the field of Big Data.

Data description

Course prices vary greatly and are largely dependent on the duration of the course and the country in which it is held.
The highest price is 35200 Euro. There still exist many couse organized by free. Fig. 18 and Table 3 presented below shows in detail the course prices in euros.

Table 3. Price (in euro)

price in Euro	number of responces
8300	1
8000	1
0	10
1499	2
10870	1
16900	1
16926	1
17000	1
2000	1
2500	2
2900	1
5200	1
20000	1
35200	1
670	1
766	1
28700	1
14500	1
10524	1
13937	1
167	1
9920	1

Density of the prices of the training courses in the Big Data (is presented below). Mean value of the price is in blue.

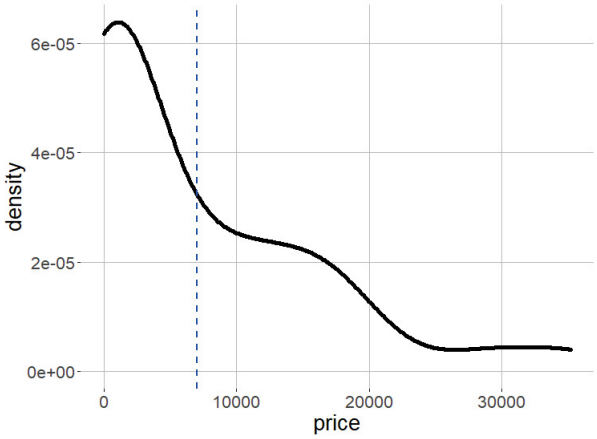

Figure 18. Density of the price Big Data courses

Discussion

For the reasons given we can see that there is big variance in course prices. The price depends very much on the area in which the course is being held and the time that is available for the course. Among Big Data field courses there could be found free courses but there are also courses for 35200 Euro.

Main conclusions:

- The highest price is 35200 Euro, but there still exist many Big Data courses organized by free.
- On average, the course fee is 6984,182 Euro.

4. Conclusions

For the reasons given, the research was conducted based on the advertising information about training courses in the field of Big Data technology and Data Science. The survey was performed on-line using google forms tools. Due to various formats and specification of advertising information in each case the data was collected by scientists based on phrase search. The analysis shows us the basic hard and soft skills and other requirements that should be taken into account when designing the training course on Big Data

REFERENCES

1. TINOCO-MARTÍNEZ, S.R., FERREIRA-MEDINA, H., CENDEJAS-VALDEZ, J.L., HERNÁNDEZ-RENDÓN, F., FLORES-MONROY, M.M., GINORI-RODRÍGUEZ, B.H.: How to Improve the Teaching of Computational Machine Learning Applied to Large-Scale Data Science: The Case of Public Universities in Mexico (2023) Lecture Notes in Networks and Systems, 544 LNNS, pp. 1-15. DOI: 10.1007/978-3-031-16075-2_1.

2. HAENLEIN, M., KAPLAN, A.: A brief history of artificial intelligence: On the past, present, and future of artificial intelligence (2019) California Management Review, 61 (4), pp. 5-14. DOI: 10.1177/0008125619864925.
3. DENG, L., YU, D.: Deep learning: Methods and applications (2013) Foundations and Trends in Signal Processing, 7 (3-4), pp. 197-387. DOI: 10.1561/2000000039.
4. HUANG, L., MA, K.-S. Introducing machine learning to first year undergraduate engineering students through an authentic and active learning lab ware (2018) 2018 IEEE Frontiers in Education Conference (FIE), pp. 1-4.
5. Webpage: *https://imp.i384100.net/zavqGm* (Last accessed 16-10-2022).
6. Webpage: *http://datacamp.pxf.io/e4NgD6* (Last accessed 16-10-2022).
7. Webpage: *https://click.linksynergy.com/deeplink?id=6m9Z7SVq*Tw&mid=42536&murl=https%3A%2F%2Fwww.edureka.co%2Fbig-data-hadoop-training-certification%3FqId%3D2eb794c4c4a8faeee4c71a45b0c54d75%26index_name%3Dprod_search_results_courses%26objId%3D888%26objPos%3D1&LSNSUBSITE=LSNSUBSITE* (Last accessed 16-10-2022).
8. Webpage: *https://www.awin1.com/cread.php?awinmid=6798&awinaffid=778131&clickref=&ued=https%3A%2F%2Fwww.edx.org%2Fcourse%2Fknowledge-management-and-big-data-in-business* (Last accessed 16-10-2022).
9. Webpage: *https://www.experfy.com/training/courses/big-data-what-every-manager-needs-to-know* (Last accessed 16-10-2022).
10. Webpage: *https://intellipaat.com/post-graduate-certification-big-data-analytics/* (Last accessed 16-10-2022).
11. Webpage: *https://linkedin-learning.pxf.io/jWWWvP* (Last accessed 16-10-2022).
12. Webpage: *https://mindmajix.com/big-data-on-aws-training* (Last accessed 16-10-2022).
13. Webpage: *http://pluralsight.pxf.io/KRMv9* (Last accessed 16-10-2022).
14. Webpage: *https://www.simplilearn.com/big-data-engineer-masters-certification-training-course?referrer=search&tag=big%20data* (Last accessed 16-10-2022).
15. Webpage: *https://click.linksynergy.com/deeplink?id=6m9Z7SVq*Tw&mid=39197&murl=https%3A%2F%2Fwww.udemy.com%2Fcourse%2Fspark-and-python-for-big-data-with-pyspark%2F* (Last accessed 16-10-2022).
16. Webpage: *https://imp.i115008.net/n111Gx* (Last accessed 16-10-2022).