Aleksandra KŁOS-WITKOWSKA[1], Kateryna MERKULOVA[2],
Tetiana KOVALIUK[2], Yelyzaveta ZHABSKA[2], Sergii DOTSENKO[2],
Andriy SHATYRKO[2], Vasyl MARTSENYUK[1], Georgi DIMITROV[3],
Eugenia KOVATCHEVA[3], Dejan RANCIC[4], Oleksii BYCHKOV[2]

KSZTAŁCENIE BIG DATA W OPARCIU O KOMPETENCJE: WYMAGANIA DLA INNOWACYJNEGO KURSU

Streszczenie: Praca obejmuje analizę wymagań i cech Big Data, które powinniśmy zbadać, aby zaproponować ramy innowacyjnego kursu Big Data. Wymagania są formułowane za pomocą odpowiednich umiejętności twardych i miękkich. Tematy innowacyjnego kursu są mapowane przez odpowiednie kompetencje.

Słowa kluczowe: Big Data, innowacyjny kurs, umiejętności twarde, umiejętności miękkie, kompetencje, iBIGworld

COMPETENCE-BASED STUDY OF BIG DATA: REQUIREMENTS OF AN INNOVATIVE COURSE

Summary: The work includes the analysis of the requirements and features of Big Data that we should research in order to offer the framework of the innovative course on Big Data. The requirements are formulated with the help of corresponding hard and soft skills. Topics of the innovative course are mapped by the corresponding competencies.

Keywords: Big Data, innovative course, hard skills, soft skills, competencies, iBIGworld

1. Introduction

The development of innovative training course which is based on competence-based learning environment is under consideration in the series of works [1-6]. The given work is devoted to setting the requirements for the Big Data training course by focusing on mapping competencies into corresponding topics of the content. The stage

[1]Department of Computer Science and Automatics, University of Bielsko-Biala, Poland: (awitkowska, vmartsenyuk)@ath.bielsko.pl

[2]Taras Shevchenko National University of Kyiv, Kiev, Ukraine: oleksiibychkov@knu.ua

[3]University of Library Studies and Information Technologies, Sofia, Bulgaria: g.dimitrov@unibit.bg, e.kovatcheva@unibit.bg

[4] University of Niš, Nis, Serbia: dejan.rancic@elfak.ni.ac.rs

of the iBIGworld project entitled by "Setup the Big Data Requirements" (output O2) is based on the intellectual output O1, namely, "DataBased with good practices Big Data case: Data Collection and Analysis". This output includes the analysis of the requirements and features of Big Data that we should research in order to offer the framework of the innovative course on Big Data. Namely:

- **Data Processing**, which involves the collection and organization of raw data to produce meaning. Data mining allows users to extract and analyze data from different perspectives and summarize it into actionable insights. It is especially useful on large unstructured data sets collected over a period of time.
- **Predictive Applications**. Identity management also deals with issues including how users gain an identity with access, protection of those identities and support for other system protections such as network protocols and passwords. It determines whether a user has access to a system and the level of access that user has permission to utilize.
- **Analytics**. Big Data analytics tools offer a variety of analytics packages and modules to give users options. Risk analytics, for example, is the study of the uncertainty surrounding any given action. It can be used in combination with forecasting to minimize the negative impacts of future events. Risk analytics allow users to mitigate these risks by clearly defining and understanding their organization's tolerance for and exposure to risk. Decision management involves the following decision making processes:
  a) Text analytics is the process of examining text that was written about or by customers.
  b) Content analysis is very similar to text analysis but includes the analysis of all formats of documentation including audio, video, pictures, etc.
  c) Statistical analytics collects and analyzes data sets composed of numbers. The goal is to draw a sample from the total data that is representative of a total population.
- **Reporting Features.** Reporting functions keep users on top of their business. Real-time reporting gathers minute-by-minute data and relays it to you, typically in an intuitive dashboard format.
- **Security Features.** Keeping the system safe is crucial to a successful business. Big Data analytics tools should offer security features to ensure security and safety.
- **Technologies Support**. Analytics software should support a variety of technology and tasks that may be useful to every user.

The purpose of the work is to develop the methodologies/guidelines as a part of the research within O2 in connection with the objectives of project iBIGworld, that were aimed at identifying soft and professional competencies, learning outcomes that shape competencies, and educational topics that shape learning outcomes. The information presented here can be used to form educational programs for Big Data courses.

2. Syllabus development methodology

A survey of stakeholders in seven nominations was conducted. According to the results of the survey, general and professional competencies were identified. which must be formed in students in the learning process. The competencies that received

the most votes of the surveyed stakeholders were suggested for consideration when compiling the syllabus.

Table 1 presents soft skills. Competencies rated from 431 to 200 were selected for inclusion in the curricula.

Table 1. Results of a stakeholder survey on soft skills

#	Competency	Total	Percentage
1	Ability to work in a team	431	68.30
2	Ability to plan and manage time	427	67.67
3	Ability to communicate in a second (foreign) language	415	65.77
4	Capacity to learn and stay up-to-date with learning	353	55.94
5	Ability to identify, propose and resolve problems	329	52.14
6	Ability to apply knowledge in practical situations	306	48.49
7	Capacity to generate new ideas (creativity)	304	48.18
8	Ability to be critical and self-critical	267	42.31
9	Ability to search, process and analyze information from a variety of sources	262	41.52
10	Ability to work autonomously	243	38.51
11	Ability to make reasoned decisions	234	37.08
12	Ability to use information and communication technologies	232	36.77
13	Commitment to tasks and responsibilities	230	36.45
14	Ability to design and manage projects	222	35.18
15	Ability to adapt to and act in new situations and cope under pressure	204	32.33
16	Ability to communicate both orally and through the written word in first language	193	30.59
17	Knowledge and understanding of the subject area and understanding of the profession	192	30.43
18	Ability to communicate with people who are not experts in the subject area	183	29.00
19	Ability to work in an international context	173	27.42
20	Ability for abstract and analytical thinking, and synthesis of ideas	168	26.62
21	Ability to interact with others in a constructive manner, even when dealing with difficult issues	166	26.31
22	Ability to motivate people and move toward common goals	163	25.83
23	Ability to evaluate and maintain the quality of work produced	159	25.20
24	Commitment to health, well-being and safety	114	18.07
25	Ability to act on the basis of ethical reasoning	110	17.43
26	Ability to undertake research at an appropriate level	105	16.64
27	Ability to interact constructively with others regardless of background and culture and respecting diversity	90	14.26

28	Commitment to conservation of the environment	81	12.84
29	Ability to act with social responsibility and civic awareness	80	12.68
30	Ability to take the initiative and to foster the spirit of entrepreneurship and intellectual curiosity	75	11.89
31	Ability to show awareness of equal opportunities and gender issues	63	9.98
21	Ability to interact with others in a constructive manner, even when dealing with difficult issues	166	26.31
22	Ability to motivate people and move toward common goals	163	25.83
23	Ability to evaluate and maintain the quality of work produced	159	25.20
24	Commitment to health, well-being and safety	114	18.07
25	Ability to act on the basis of ethical reasoning	110	17.43
26	Ability to undertake research at an appropriate level	105	16.64
27	Ability to interact constructively with others regardless of background and culture and respecting diversity	90	14.26
28	Commitment to conservation of the environment	81	12.84
29	Ability to act with social responsibility and civic awareness	80	12.68
30	Ability to take the initiative and to foster the spirit of entrepreneurship and intellectual curiosity	75	11.89
31	Ability to show awareness of equal opportunities and gender issues	63	9.98

Table 2 presents the professional competencies proposed to be included in the curricula. When selecting professional competencies, we were guided primarily by the requirements of employers and the labor market. The requirements of teachers and graduates were taken into account to develop the motivation of teachers and students as participants in the educational process. The documents [7-9] were used to determine learning topics and learning outcomes

Table 2. Results of a stakeholder survey on professional skills

#	Competencies	Respondent
1	Ability to process large volumes of data using hierarchical storage, hashing and filtering.	Labor market
2	Ability to select the efficient algorithm to Big Data, which takes under consideration its scale.	Labor market
3	Ability to model, analyze, and evaluate organization's business processes.	Labor market
4	Ability to select appropriate sampling and filtering method for given Big Data analysed case.	Teacher
5	Ability to propose concepts, models, create and adapt methods and tools for professional activities using resources from others sciences.	Teacher
6	Effectively use variety of data analytics techniques (Machine Learning, Data Mining, Prescriptive and Predictive Analytics)	Teacher
7	Design, build, operate relational and nonrelational databases (SQL and NoSQL)	Teacher

8	Apply quantitative techniques (statistics, time series analysis, optimization, and prediction)	Teacher
9	Ability to tackle with concurrency / parallelism problems of Big Data scale	Alumni
10	Using wide range of Big Data analytics platforms	Alumni
11	Develop and operate large scale data storage (e.g., Data Lakes, Hadoop and others)	Alumni
12	Process large dataset (i.e., ETL, OLTP, OLAP)	Alumni
13	Ability to design relational databases	Employer
14	Ability to design local, network-based, and web information systems	Employer
15	Ability to carry out system reengineering and to analyze legacy databases	Employer
16	Ability to plan and carry out comprehensive analysis of datasets that are diverse, complex, stored in numerous databases, and very large (Big Data skills)	Employer
17	Ability to implement software components of information systems following given design documents and using various programming languages and tools	Employer
18	Ability to implement cloud computing based solutions	Employer
19	Ability to implement and administer databases and warehousing	Employer
20	Ability to plan, perform, and manage manual or automated testing	Employer
21	Ability to innovate and modify methods and approaches used in the organization;	Employer
22	Ability to perform simulations and experiments, and to critically analyze their results using statistical methods;	Employer
23	Ability to carry out feasibility studies on new technologies, methods, and standards that could be of use to the organization;	Employer

3. Results on Soft Skills

Our results presents soft skills competencies, learning outcomes and course topics that form the relevant competencies , also it is present professional competencies, learning outcomes and course topics that form the relevant competencies.

3.1. Ability to work in a team

For competency name ability to work in a team, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are principles of teamwork, team values, basics of conflictology and IT project management methodologies, RMVOK standards, software tools for IT project management .Learning outcomes for skills are build connections and relationships with people, take into account the opinion of colleagues, understand other people, express confidence in the team, admit their mistakes, avoid and prevent conflicts, restrain personal ambitions. to select and prepare information and tasks for the project team, set goals, formulate tasks for the implementation of projects and programs, plan communications in the team and with customers, be responsible for the results of the team, be responsible as a leader to the team.

For achieving ability to work in a team competencies and learning outcomes listed below study topics have been created:

1. IT project management,
1.1 Basics of conflictology,
1.2 Team decision making,
1.3 Technologies of interpersonal and group communication in business interaction.

3.2. Ability to plan and manage time

For competency name ability to plan and manage time, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: principles of time management; problems and methods of optimizing time in various areas of professional activity (Pareto, Eisenhower, ABC); time management processes (analysis, modeling of strategies, formulation of goals, work planning, implementation, and control of achievement. Learning outcomes for skills are: plan and set goals for SMART; set priorities and deadlines for each task; analyze the cost of time resources; make lists of tasks; self-organize and delegate; eliminate unnecessary tasks.

For achieving ability to work in a team competencies and learning outcomes listed below study topics have been created:

2. Time management fundamentals,
2.1. The concept of good time management,
2.2. Analysis of your time management,
2.3. Compare with your goals,
2.4. Effective prioritization of your goals,
2.5. Planning, implementation and control of your productivity,
2.6. Improving the quality of work.

3.3. Ability to communicate in the second (foreign) language

For competency name ability to communicate in the second (foreign) language, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: lexical, grammatical, stylistic features of foreign vocabulary, terminology in the field of Data Science and Big Data, grammatical structures for understanding and producing orally and in writing foreign texts in the professional sphere.

Learning outcomes for skills are: communicate in foreign languages at a professional level, develop foreign language documentation for systems and products in the field of Data Science and Big Data, read, understand and apply technical documentation in foreign languages in professional activities, build communication in oral and written form in foreign languages, based on the goals and situation of communication

 For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

3. Foreign language of professional orientation

3.4. Capacity to learn and stay up-to-date with learning

For competency name capacity to learn and stay up-to-date with learning, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are:

ways and methods of teaching, methods of self-education, basics of scientific and research activities, methods of search, collection, analysis and processing of information.

Learning outcomes for skills are: evaluate the subject of educational activities, determine the general purpose and specific tasks, choose adequate means to solve them to achieve results, exercise the necessary self-control, use reference literature and technical documentation, to develop and apply the creative abilities in creative activity, organize the workplace, plan working hours ,use media to obtain, process and create up-to-date information in the field of Data Science and Big Data in the form of documents, abstracts, reports, articles, reviews, interviews, etc. For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

4. Fundamentals of scientific and research activities,
4.1. Research methodology,
4.2. Characteristics of scientific research, its main stages and participants,
4.3. Language and style of scientific research,
4.4. Ethics of scientific research. Scientometry and plagiarism

3.5. Ability to identify, propose and resolve problems

For competency name ability to identify, propose and resolve problems, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: the stages of the creative process, the role of the correct formulation of goals and objectives for their achievement in professional, creative potential of man, the mechanism of genesis and development of knowledge, methods of generating ideas, understanding of creativity as a universal process of generating innovative ideas.

Learning outcomes for skills are: find creative solutions or answers to clearly defined specific and abstract problems based on the identification and application of data, show curiosity, ability to think and be inspired by new ideas, implement ideas and problems, combine and experiment in the process of solving the problem, show independence and responsibility for generating new ideas and decision-making in the field of Data Science and Big Data.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

5. Solving problems,
5.1. Causal Analysis,
5.2. PEST analysis,
5.3. SWOT analysis,
5.4. Five whys,
5.5. The Reframing Matrix,
5.6. Random input,
5.7. Decision trees.

3.6. Ability to apply knowledge in practical situations

For competency name ability to apply knowledge in practical situations, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: Evaluation of Big Data, according to it volume, variety, velocity and veracity ; The understanding of specifics in parallel computation models; Strategies backup,

migration and compression for data-centric systems .Learning outcomes for skills are: Removing of unwanted redundancy from Big Data; Formulate well-tuned algorithms within a parallel programming paradigm; Using of web services for the support of data analytics and machine learning in the context of Big Data

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

   6.1 Measurement in the context of Big Data;
   6.2 Classification of parallel computation models;
   6.3 Distributed systems;
   6.4 Data backup,
   6.5 Paradigm MapReduce;
   6.6. Programming environments to support Big Data applications.

3.7. Capacity to generate new ideas (creativity)

For competency name capacity to generate new ideas (creativity), learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: Evaluation of Big Data, according to it volume, variety, velocity and veracity; The understanding of specifics in parallel computation models; Knowledge of strategies backup, migration and compression for data-centric systems.

Learning outcomes for skills are: Making of data sketch; Evaluation of the complexity of a range of commonly used algorithms. For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

   7.1. Measurement in the context of Big Data,
   7.2. Classification of parallel computation models,
   7.3.Data backup,
   7.4.The efficiency of application programs,
   7.5.Computational complexity and efficiency of algorithms.

3.8. Ability to be critical and self-critical

For competency name ability to be critical and self-critical, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: understanding of risks associated with Cloud Computing, the understanding of specifics in parallel computation models.

Learning outcomes for skills are: The comparing of different parallel computing architectures: multicore, grid computing, GPUs, shared memory, distributed memory, symmetric multiprocessing, vector processing; The fitting of parallel algorithm according to available hardware architecture; Formulate well-tuned algorithms within a parallel programming paradigm; Making of data sketch.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

   8.1. Measurement in the context of Big Data;
   8.2. Concept of data storages;
   8.3. Big Data Computing Architectures;
   8.4. Classification of parallel computation models;
   8.5. Selection of parallel algorithms according to available architecture;
   8.6. Computational complexity and efficiency of algorithms.

3.9. Ability to search, process and analyze information from a variety of sources

For competency name ability to search, process and analyze information from a variety of sources, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge is the understanding of specifics in parallel computation models.

Learning outcomes for skills are: Evaluation of Big Data, according to it volume, variety, velocity and veracity; Coordinating of tasks with increasing numbers of agents / processes and accommodating scalability ; Generalization of requirements to data-centric high-performance computing; The fitting of parallel algorithm according to available hardware architecture ; Evaluation of the complexity of a range of commonly used algorithms.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

   9.1. Measurement in the context of Big Data,
   9.2. Concept of data storages,
   9.3. Big Data Computing Architectures,
   9.4. Classification of parallel computation models,
   9.5. Selection of parallel algorithms according to available architecture,
   9.6. Computational complexity and efficiency of algorithms.

3.10. Ability to work autonomously achieving ability to work in a team

For competency name ability to work autonomously achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

   10.1 Big Data Computing Architectures,
   10.2 Classification of parallel computation models,
   10.3 Data backup,
   10.4 The efficiency of application programs,
   10.5 The advantages of Cloud Computing,
   10.6 Security issues for cloud including cloud computing, cloud storage and virtual machine.

3.11. Ability to make reasoned decisions

For competency name ability to make reasoned decisions, learning outcomes are created by analyze, evaluate and select tools and computing tools, technologies, algorithmic. and software solutions to solve Big Data problems.

 For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

   11.1. Technologies for Handling Big Data (Big Data is primarily characterized by Hadoop. This module cover topics such as Introduction to Hadoop, functioning of Hadoop, Cloud computing (features, advantages, applications) etc.),
   11.2. Big Data regulatory frameworks; Knowledge of state documentation of Big Data governance,
   11.3. Security and Privacy in Big Data Technology Standards for security and privacy-preserving of Big Data.

3.12. Ability to use information and communication technologies

For competency name ability to use information and communication technologies, learning outcomes are created by use information and communication technologies in communication, exchange, gathering, analysis and processing of information.
For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

12.1. Big Data Analytics in the Cloud (Use of Big Query and Cloud Storage through using either legacy or standard SQL queries; Use of Big Query ML to train advanced machine learning models with data inside Big Query; Use of Big Query GIS to ingest, process, and analyze geospatial data),
12.2. Big Data Software Testing Create Big Data testing strategy with the use of testing tools: Hadoop, HPCC, Cloudera, Cassandra, Storm,
12.3. Databases and Data Warehouses (It covers all about databases, polygot persistence and their related introductory knowledge).

3.13. Commitment to tasks and responsibilities

For competency name commitment to tasks and responsibilities, learning outcomes are created by knowing the code of professional ethics, understand the social significance and cultural aspects of Big Data and adhere to them in professional activities.
For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

13.1. Introduction to Big Data Analytics (Introduction into the concepts of Big Data analysis, the role of a Big Data Analyst, and the tools that are used to perform daily functions),
13.2. Using Big Data in Businesses (It focus on the application perspective of Big Data covering topics such as using Big Data in marketing, analytics, retail, hospitality, consumer good, defense etc),
13.3. Big Data Analytics Practices that help organizations to implement Big Data strategy to meet analytics needs and produce the expected business benefits.

3.14. Ability to design and manage projects

For competency name ability to design and manage projects learning outcomes are created by knowing and applying Big Data project design and management frameworks. For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

14.1. Programming Models for Big Data,
14.2. The Hadoop Distributed File System: A Storage System for Big Data,
14.3. Big Data analytics implementation in organizations (Kotter's change model and the Six Sigma model frameworks for implementing Big Data analytics).

3.15. Ability to adapt to and act in new situations and cope under pressure

For competency name ability to adapt to and act in new situations and cope under pressure learning outcomes are created by ability to find and troubleshoot the problem in Big Data technology, collaborate with the team.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

15.1. Big Data Disaster Recovery Technologies (NoSQL and Hadoop Backup and Recovery),

15.2. Big Data Software: High-Performance Modelling and Simulation (M&S-based approaches of Big Data access in distributed environments),

15.3. Methods to Inspire Innovation Within Organization to use Big Data technologies.

4. Results on hard skills

Our results presents also professional skills created by competencies, learning outcomes and course topics.

4.1. Ability to process large volumes of data using hierarchical storage, hashing and filtering

For competency name ability to process large volumes of data using hierarchical storage, hashing and filtering, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: approaches to storing large amounts of data, including storage on different devices, hierarchies of Big Data storage ,algorithms and techniques of hashing, sampling, filtering and their limitations in the process of Big Data processing, data sampling and filtering criteria.

Learning outcomes for skills are: use storage hierarchies to work with Big Data, process Big Data using hierarchical storage, criticaly analyze the application of different approaches to data filtering, apply hashing algorithms to search data in a database, analyze big files, and manage Big Data.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

1. Distributed Data Storage,

1.1. Distributed Storage System (DSS) as advanced form of the "Software-Defined Storage" concept ,

1.2. types of storage (block, file, and object),

1.3. Flexibility, speed and cost of distributed storage system,

2. Hierarchical storage management

2.1.Implementation,

2.2. Tiered storage, Data migration policy

4.2. Ability to select the efficient algorithm to Big Data, which takes under consideration its scale.

For competency name ability to select the efficient algorithm to Big Data, which takes under consideration its scale, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: methods and algorithms of machine learning and deep learning on Big Data sets; technologies for finding patterns and predicting events using large arrays of sensory data; deep learning methods for analyzing large arrays of multimedia data; algorithms for parallel processing of Big Data; high-performance computing technologies.

Learning outcomes for skills are: use a distributed parallel computing model in MapReduce computer clusters to process large data; use NoSQL technology for complex and flexible queries specific to Big Data; use Hadoop utilities, libraries, and frameworks to develop and run distributed applications running on clusters of hundreds and thousands of nodes, store data in Hadoop using the HDFS (Hadoop Distributed File System) file system with access to data on multiple cluster nodes; use the programming language R for analysis and statistical processing of large data.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

  3. Algorithms for Big Data

  3.1 Big Data algorithms for data ingest, pre-processing, and visualization;

  3.2. Big Data algorithms for large scale data processing;

  3.3. Big Data Analytics;

  3.4. Big data processing technologies;

  3.5. Machine learning and deep learning;

  3.6 Association rule learning,

  3.7 Numeric prediction, 3.8.Parallel processing of Big Data

  3.9 High-performance computing technologies, 3.10   Data mining doe Big Data, 3.11. Statistical methods

4.3. Ability to model, analyze, and evaluate organization's business processes

For competency name ability to model, analyze, and evaluate organization's business processes, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: knowing and understanding the essence of the process approach to the management of organizations; awareness of the key process-based management concepts; methodologies and strategies of performing practical projects involving organizational changes intended to design and enhance business processes, factors to enhance the progress of the processes and provide reasons evidencing their efficacy.

Learning outcomes for skills are: identify the key and auxiliary business processes as well as their objectives and indicators; use the systems of ARIS, ADONIS, and etc. with a view to modelling and analyzing the length of performance of the processes as well as their standard and the costs involved; collect information concerning the progress of the processes.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

  4 Business Process Analysis and Modeling,

  4.1. Business Process Modeling (BPM) Concepts,

  4.2. Enterprise Process Context Modeling,

  4.3 Framing the Process,

  4.4. Process Mapping,

  4.5. Refining Process Maps,

  4.6. BPMN notation,

  4.7 Process Maps in Requirements

  4.8 ARIS – the methods of process modeling.,

  4.9 ADONIS – the methods of process, work environment, and documents modeling, the tasks of modeling with the use of individual methods.

4.10 Business process effectiveness management, performance process management.

4.11.The concepts and solutions that support the process attitude application: outsourcing, benchmarking, reference models, computer solutions in process management introduction.

4.4. Ability to select appropriate sampling and filtering method for given Big Data analyzed case

For competency name ability to select appropriate sampling and filtering method for given Big Data analyzed case, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are variety of approaches to filtering, illustrating their use. Learning outcomes for skills are: perform sample selection for a particular application involving Big Data.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

  5.1. The role of sampling and filtering in the processing of Big Data,

  5.2 Benefits of sampling/filtering,

  5.3. Criteria to be used in guiding typical sample selection.

4.5. Effectively use variety of data analytics techniques (Machine Learning, Data Mining, Prescriptive and Predictive Analytics)

For competency name effectively use variety of data analytics techniques (Machine Learning, Data Mining, Prescriptive and Predictive Analytics), learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: methods of data mining; crowdsourcing method, data fusion and integration methods for integrating disparate data from different sources for deep analysis (digital signal processing, natural language processing, including tone analysis, etc.), mmethods of machine learning with and without a teacher, deep learning for the tasks of forecasting, analysis, identification, recognition,

Learning outcomes for skills are: use Data Mining, TextMining, WebMining technologies for data mining, solve professional problems using methods of classification, forecasting, cluster analysis, search for associative rules, etc., solve optimization and modeling problems using artificial neural networks, network analysis, and genetic algorithms.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

  6. Data mining,

  6.1. Classification, clustering, regression analysis,

  6.2. Forecasting,

  6.3. Learning associative rules,

  6.4. Data visualization,

  6.5. Detection and analysis of deviations,

  6.6. Evaluation, analysis of relationships.

  7. Methods and algorithms of machine learning:

  7.1. Neural networks,

  7.2. Decision trees,

  7.3. Symbolic rules,

7.4. Methods of the nearest neighbor and k-nearest neighbor,

7.5. Method of support vector machines,

7.6. Hierarchical and non-hierarchical methods of cluster analysis;

7.7. Methods of searching for associative rules, in particular, the Apriori algorithm; limited search method,

7.8. Evolutionary programming, and genetic algorithms,

7.9. Data visualization methods.

8. Neural network technologies,

8.1. Feed forward neural networks

8.2. Multilayer neural network,

8.3. Recurrent neural networks,

8.4. Convolutional neural networks

8.5. Markov chains.

4.6. Design, build, operate relational and nonrelational databases (SQL and NoSQL)

For competency name design, build, operate relational and nonrelational databases (SQL and NoSQL), learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: modern theories of organization of databases and knowledge, methods and technologies of their development, approaches to storing huge amounts of data, including storage on different devices, storage hierarchies, providing clear, consistent and representative data, protection and support, concepts of automatic scaling and serverless computing.

Learning outcomes for skills are: work with relational and non-relational databases (MySQL, MSSQL, PostgreSQL, MongoDB, SQL Server, Oracle, HP Vertica, Amazon Redshift, etc.), use cloud platforms for Big Data solutions (Amazon Web Services, Google Cloud Platform, Microsoft Azure, and other similar solutions from large PaaS / IaaS providers); configure infrastructure for corporate storage, ETL systems, internal databases, and third-party sources (mail, CRM, ERP, and other application systems); organize automated data collection from various sources into single centralized storage (Data Warehouse) or Data Lake, move and store large data sets; configure, integrate and create data showcases for analysts and researchers.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

10. Algorithms and data structures,

10.1. Data search algorithms and technologies,

10.2. Data sorting algorithms,

10.3. Relational and non-relational databases,

10.4. Installation SQL Server, Analysis Server (Cube), Analysis Server (Tabular), and MongoDB,

10.5 creation databases in all installations,

10.6 Relational data modeling (SQL Server),

10.11. Non-relational data modeling (cubes, tabular, and document databases),

10.12. Cloud services and database implementation.

11. Distributed Data Storage,

11.1. Data Warehouse and Data Lake

11.2. Create pipelines for regular and continuous data preparation (CI / CD pipelines);

11.3. Develop storage systems with appropriate strategies (backup, migration, compression) to ensure scalability, usability, efficiency, and security,

11.4. Load balancing and automatic scaling and data spaces.

4.7. Apply quantitative techniques (statistics, time series analysis, optimization, and prediction)

For competency name apply quantitative techniques (statistics, time series analysis, optimization, and prediction), learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are methods of statistical data analysis: analysis of variance, correlation and regression analysis, covariance analysis, component and factor analysis, cluster and discriminant analysis, time series analysis and forecasting Learning outcomes for skill is: use methods and algorithms of statistical data analysis in professional activities.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

12.1.Statistical Modeling for Data Science Applications,

12.2. Statistical Analysis.

4.8. Ability to tackle with concurrency / parallelism problems of Big Data scale

For competency name ability to tackle with concurrency / parallelism problems of Big Data scale, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: understanding of specifics in parallel computation models The fitting of parallel algorithm according to available hardware architecture, formulate well-tuned algorithms within a parallel programming paradigm, identify the overheads and computational complexity associated with parallelism in particular algorithms . Learning outcomes for skills are: removing of unwanted redundancy from Big Data, compare different parallel computing architectures: multicore, grid computing, GPUs, shared memory, distributed memory, symmetric multiprocessing, vector processing.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

13.1. Classification of parallel computation models,

13.2. Distributed systems,

13.3. Selection of parallel algorithms according to available architecture,

13.4. Paradigm MapReduce,

13.5. Complexity of concurrent algorithms,

13.6. The efficiency of application programs.

4.9. Using wide range of Big Data analytics platforms

For competency name using wide range of Big Data analytics platforms, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge is: the characteristics, capabilities, features of Big Data analytics platform.

Learning outcomes for skills are: use the following platforms for professional activities; Microsoft Azure, Cloudera, Sisense, Collibra, Tableau, MapR, Qualtrics, Oracle, MongoDB, Datameer.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

14.1. Big Data Computing Architectures,

14.2. Distributed systems,

14.3. Storage hierarchies of data,

14.4. The advantages of Cloud Computing,

14.5. Security issues for cloud including cloud computing, cloud storage and virtual machines.

4.10. Develop and operate large scale data storage

For competency name develop and operate large scale data storage (e.g., Data Lakes, Hadoop and others) , learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: Hadoop's Core Components; The Hadoop Distributed File System; Hadoop Common collection of standard utilities and libraries that support other modules; Security on Hadoop and Implementing Data Lakes. Learning outcomes for skills are: use Hadoop for complex calculations and difficult simulations, use Hadoop to integrate different systems into a single data lake foundation; Execute secure data transitions using Hadoop security tools.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

15.1. Measurement in the context of Big Data,

15.2. Concept of data storages,

15.3. Storage hierarchies of data,

15.4. Data backup

15.5. Computational complexity and efficiency of algorithms,

15.6. Programming environments to support Big Data applications.

4.11. Process large dataset

For competency name process large dataset (i.e., ETL, OLTP, OLAP), learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: difference between OLTP and OLAP, ETL (extract, transform, load) as the backbone of data processing for modern enterprises, aarchitecture and operating principle of ELT-system. Learning outcomes for skills are: perform ETL process for movement of information (data flow) from source to destination through the staging area, load ETL raw data, validate data, configure compliance, aggregate data, upload to the target system.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

16.1. Measurement in the context of Big Data,

16.2. Concept of data storages,

16.3. Big Data Computing Architectures,

16.4. Storage hierarchies of data,

16.5. Data backup,

16.6. The efficiency of application programs.

4.12. Ability to design relational databases

For competency name ability to design relational databases learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: design a relational database, normalization DB and each of the three normal forms
Learning outcomes for skills are: designing a database from nothing, identify tables, and how to create relationships.
For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

    17.1. Measurement in the context of Big Data,
    17.2. Concept of data storages,
    17.3. Big Data Computing Architectures,
    17.4. Distributed systems,
    17.5. Data backup the efficiency of application programs.

4.13. Ability to design local, network-based, and web information systems

For competency name ability to design local, network-based, and web information systems, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are: generalization of requirements to data-centric high-performance computing, coordinating of tasks with increasing numbers of agents / processes and accommodating scalability, knowledge of strategies backup, migration and compression for data-centric systems. Learning outcomes for skills are: making of data sketch ,knowledge of data center architecture, approaches to supporting Cloud Computing, cloud Services for Big Data applications , using of web services for the support of data analytics and machine learning in the context of Big Data.
For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

    18.1. Concept of data storages,
    18.2. Big Data Computing Architectures,
    18.3. Data backup,
    18.4. The efficiency of application programs,
    18.5. The advantages of Cloud Computing,
    18.6. Programming environments to support Big Data applications.

4.14. Ability to carry out system reengineering and to analyze legacy databases

For competency name ability to carry out system reengineering and to analyze legacy databases, learning outcomes are created by knowledge and skills. Learning outcomes for knowledge are :the understanding of methods used to hashing, filtering, sampling of data , coordinating of tasks with increasing numbers of agents / processes and accommodating scalability, the comparing of different parallel computing architectures: multicore, grid computing, GPUs, shared memory, distributed memory, symmetric multiprocessing, vector processing, identify the overheads and computational complexity associated with parallelism in particular algorithms, data center architecture, approaches to supporting Cloud Computing, Cloud Services for Big Data applications .Learning outcomes for skill is: evaluation of Big Data, according to it volume, variety, velocity and veracity.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

19.1. Measurement in the context of Big Data,

19.2. Concept of data storages,

19.3. Big Data Computing Architectures,

19.4. Storage hierarchies of data,

19.5. Complexity of concurrent algorithms,

19.6. The advantages of Cloud Computing.

4.15. Ability to plan and carry out comprehensive analysis of datasets that are diverse, complex, stored in numerous databases, and very large

For competency ability to plan and carry out comprehensive analysis of datasets that are diverse, complex, stored in numerous databases, and very large (Big Data skills), learning outcomes are created by knowledge and skills.

Learning outcomes for knowledge are: data center architecture, approaches to supporting Cloud Computing, Cloud Services for Big Data applications, the fitting of parallel algorithm according to available hardware architecture. Learning outcomes for skills are: using of web services for the support of data analytics and machine learning in the context of Big Data, removing of unwanted redundancy from Big Data, making of data sketch, evaluation of the complexity of a range of commonly used algorithms.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

20.1. Distributed systems,

20.2. Selection of parallel algorithms according to available architecture,

20.3. The efficiency of application programs,

20.4. The advantages of Cloud Computing,

20.5. Computational complexity and efficiency of algorithms,

20.6. Programming environments to support Big Data applications.

4.16. Ability to implement software components of information systems following given design documents and using various programming languages and tools

For competency name ability to implement software components of information systems following given design documents and using various programming languages and tools, learning outcomes are created by knowledge and skills. Learning outcomes is knowing, understanding, analyzing and select algorithms and software for Big Data analysis and governance in the process of development of software components of information systems.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

21.1. Programming Models for Big Data,

21.2. The Hadoop Distributed File System: A Storage System for Big Data,

21.3. Learn Hadoop YARN architecture. This module covers the background of YARN, advantages of YARN,

21.4. Working with YARN, backward compatibility with YARN, YARN Commands, log management etc.,

21.5. Functional Programming in Scala for Big Data Processing (Scala is suitable for working with Big Data tools like Apache Spark for distributed Big Data processing. Go for Big Data. Go-based systems are being used to integrate machine learning and parallel processing of data)

4.17. Ability to implement cloud computing based solutions

For competency name ability to implement cloud computing based solutions, learning outcomes are created by knowledge and skills. Learning outcomes are: implement distributed computing based on cloud services and technologies, parallel and distributed computing in the development and operation of distributed parallel Big Data processing systems.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

22.1. Cloud Computing: An Important Big Data Enabler,

22.2. Scalable Computing Over the Internet,

22.3. Big Data Analytics in the Cloud, use of BigQuery and Cloud Storage through using either legacy or standard SQL queries. Use of BigQuery ML to train advanced machine learning models with data inside BigQuery. Use of BigQuery GIS to ingest, process, and analyze geospatial data,

22.4. Big Data Storage with Cloud Volumes ONTAP (Cloud Volumes ONTAP is the data management platform that delivers the trusted NetApp ONTAP capabilities to Azure, GCP, and AWS).

4.18. Ability to implement cloud computing based solutions

For competency name ability to implement and administer databases and warehousing, learning outcomes are created by knowledge and skills. Learning outcomes for is: know and be able to design logical, conceptual and physical models of Big Databases and warehouses, queries to them and apply various database management systems.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

23.1. Databases and Data Warehouses (It covers all about databases, polygot persistence and their related introductory knowledge),

23.2. Using Hadoop to store data (This includes an entire module of HDFS, HBase and their respective ways to store and manage data along with their commands).

23.3. Big Data Disaster Recovery Technologies NoSQL and Hadoop Backup and Recovery,

23.4. Blockchain Storage Platforms for Big Data (Knowledge of Storj, an open-source, decentralized file storage solution that uses cryptography, sharing, and harsh tables to help store files on a peer-to-peer network, and Datum, a decentralized storage network driven by DAT, the Data Access Token).

4.19. Ability to plan, perform, and manage manual or automated testing

For competency name ability to plan, perform, and manage manual or automated testing, learning outcomes are created by knowledge and skills. Learning outcomes

is: Critically analyze Big Data datasets and implementations, taking practicality and usefulness metrics into consideration.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

24.1. Technologies for Handling Big Data (Big Data is primarily characterized by Hadoop. This module cover topics such as Introduction to Hadoop, functioning of Hadoop, Cloud computing (features, advantages, applications) etc.),

24.2. Strategies of Big Data Testing (Testing application that handles Big Data would take the skill from a whole new level and out of the box thinking. The core and important tests is based on three Scenarios: Batch, Real-Time and Interactive Data Processing Test),

24.3. Big Data Software Testing (Create Big Data testing strategy with the use of testing tools: Hadoop, HPCC, Cloudera, Cassandra, Storm).

4.20. Ability to innovate and modify methods and approaches used in the organization

For competency name ability to innovate and modify methods and approaches used in the organization, learning outcomes is: understand and apply Big Data concepts and methods to solve problems in real-world contexts.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

25.1. Methods to Inspire Innovation Within Organization to use Big Data technologies,

25.2. Using Big Data in Businesses. (It focuses on the application perspective of Big Data covering topics such as using Big Data in marketing, analytics, retail, hospitality, consumer good, defense etc),

25.3. Big Data analytics implementation in organizations Kotter's change model and the Six Sigma model frameworks for implementing Big Data analytics.

4.21. Ability to perform simulations and experiments, and to critically analyze their results using statistical methods

For competency name ability to perform simulations and experiments, and to critically analyze their results using statistical methods, learning outcomes is: apply basic knowledge of fundamental and applied mathematics, computational and probability-statistical method in the development of software for Big Data processing tasks.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

26.1. Exploring Hive. This modules introduces you with all the necessary knowledge of Hive,

26.2. Integrating R and Hadoop and Understanding Hive in Detail This module introduces you to RHadoop, ways to do text mining and related knowledge.

26.3. Big Data Software: High-Performance Modelling and Simulation M&S-based approaches of Big Data access in distributed environments.

4.22. Ability to carry out feasibility studies on new technologies, methods, and standards that could be of use to the organization

For competency name ability to carry out feasibility studies on new technologies, methods, and standards that could be of use to the organization, learning outcomes is: understand, analyze and select the necessary information and reference resources to solve professional problems, taking into account modern advances in science and technology.

For achieving ability to work in a team competencies and learning outcomes listed below study topic have been created:

27.1. Learn NoSQL Data Management (This modules covers all about NoSQL including document databases, relationships, graph databases, schema less databases, CAP Theorem etc),

27.2 Big Data regulatory frameworks. Knowledge of state documentation of Big Data governance,

27.3. Security and Privacy in Big Data Technology. Standards for security and privacy-preserving of Big Data.

5. Conclusions

The work presented the analysis of the requirements and features of Big Data that we should research in order to offer the framework of the innovative course on Big Data. The requirements were formulated with the help of corresponding hard and soft skills. Topics of the innovative course were mapped by the corresponding competencies.

REFERENCES

1. MISBAH Z., GULIKERS J., WIDHIARSO W., MULDER M.: Exploring connections between teacher interpersonal behaviour, student motivation and competency level in competence-based learning environments (2022) Learning Environments Research, 25 (3), pp. 641-661. DOI: 10.1007/s10984-021-09395-6.
2. BIEMANS H., NIEUWENHUIS L., POELL R., MULDER M., WESSELINK R.: Competence-based vet in the netherlands: Background and pitfalls (2004) Journal of Vocational Education and Training, 56 (4), pp. 523-538. DOI: 10.1080/13636820400200268.
3. 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.

4. 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.

5. 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.

6. HUANG L., MA K.-S.: Introducing machine learning to firstyear undergraduate engineering students through an authentic and active learning labware (2018) 2018 IEEE Frontiers in Education Conference (FIE), pp. 1-4.

7. E-Competence Framework.

8. File: *ACM DataScience_Final_Report2021.pdf* .

9. ACM Computing Competencies for Undergraduate Data Science Curricula 2019.pdf.

Michał KUPCZAK[1], Marcin BERNAŚ[1], Vasyl MARTSENYUK[1],
Georgi DIMITROV[2], Dejan RANCIC[3], Oleksiy BYCHKOV[4]

INTELIGENTNY HUB PRACOWNIKÓW BIG DATA

Streszczenie: W artykule przedstawiono wdrożenie portalu Smart Big Data Job Hub opracowanego na podstawie wymagań projektu IBigWorld. Głównym celem portalu jest promowanie możliwości biznesowych pomiędzy uczelniami a firmami. Jako naturalną ewolucję portalu zaproponowano dwa rozwiązania: szablonowe i autorskie. Na koniec przedstawiono statystyki użytkowania i propozycję dalszego rozwój.

Słowa kluczowe: CMS, aplikacje webowe, frontend, backend, monolit modularny

SMART BIG DATA JOB HUB

Summary: The paper presents the implementation of Smart Big Data Job Hub portal developed based on requirements of IBigWorld project. Main goals of the portal are to promote business opportunities, between universities and business. Two solutions were proposed as natural evolution of portal: templated based and author solution. Finally, the usage statistics and further development plan was presented.

Keywords: CMS, Web applications, frontend, backend, modular monolith

1. Introduction

Project "Innovations for Big Data in a Real World" (iBIGworld) brought together higher education institution (HEI), business represented by companies and civil institutions to address the need of competence and job profile of employee needed on the market. The project [1-2] brings together four partners from four European countries: Poland, Bulgaria, Ukraine, Serbia. This collaboration provided innovative solutions to develop future Big Data experts. The prepared learning framework was based on IEEE guidelines for Big Data in Machine Learning". In this context the project aimed to not only create a bridge to fill the digital skills gap in southern Europe

[1] Department of Computer Science and Automatics, University of Bielsko-Biala, Poland: mk050666@student.ath.edu.pl, (mbernas,vmartsenyuk)@ath.bielsko.pl
[2] University of Library Studies and Information Technologies, Sofia, Bulgaria: g.dimitrov@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