LAMBDA LEARNING AND CONSULTING PLATFORM

VALENTINA JANEV, DEJAN PAUNOVIĆ

University of Belgrade, Institute Mihajlo Pupin, valentina.janev@pupin.rs

EMANUEL SALLINGER

University of Oxford, Department of Computer Science, emanuel.sallinger@cs.ox.ac.uk

DAMIEN GRAUX

ADAPT SFI Centre, Trinity College Dublin, damien.graux@adaptcentre.ie

Abstract: *The potential behind the exploitation of data (Open, Linked and Big) to boost economies and growth is in the focus of many EU initiatives, the most recent of which is the European Strategy for Data. Many students and professionals are taking the online route to acquire knowledge on emerging technologies for Big Data. Therefore, in this paper, we introduce the LAMBDA Learning and Consulting platform that contains over 30 lectures related to topics such as challenges in processing Big Data, semantic and knowledge graphs-based tools for Big Data, Big Data architectures, Smart Data Analytics, Best Practices and Use Cases from different industries. The paper discusses the technical elements of the platform and the possibilities for adoption in different settings including vocational training.*

Keywords: *E-Learning, Big Data, Analytics, Open education, Platform, Lectures*

1. INTRODUCTION

The topics of Big Data, Linked Data, Open Data, Semantic technologies and Smart Analytics have spawned a tremendous amount of attention among scientists, industry leaders and decision makers in Europe, in the last decade. However, despite the strong scientific output and high public investment, Europe still lags behind the United States and China in high-tech industries. For instance, with the emergence of Big Data, the last decade also witnessed a technology boost for AI-driven technologies. From 2013 through 2016, external investment in AI technologies had a compound annual growth rate of almost 40 percent [1]. So far, external investment remains highly concentrated geographically, dominated by a few technology hubs in the United States and China, with Europe lagging far behind. In order to realize the EU vision of attractive, secure and dynamic data economy, and to implement the European Strategy for Data [2], which aims at creating a single market for data that will ensure Europe's global competitiveness and data sovereignty, different initiatives have been initiated such as

- investing in next generation tools and infrastructures to store and process data (see for instance the European Open Science Cloud) [3];
- pooling European data in key sectors, with common and interoperable data spaces, see for instance BDVA i-spaces [4].

With the aim of aligning the activities of research organizations from Serbia and the region with existing efforts across Europe, in July 2018 the Serbian R&D Institute Mihajlo Pupin started with the implementation of the EU project LAMBDA – **L**earning, **A**pplying, **M**ultiplying **B**ig **D**ata **A**nalytics (H2020, GA. 809965). In the last two years, the partners (Institute Mihajlo Pupin, Serbia; Fraunhofer Institute for Intelligent Analysis and Information Systems, Germany; Institute for Computer Science - University of Bonn, Germany; Department of Computer Science - University of Oxford, UK) studied many aspects related to Big Data and semantic technologies and proposed a training approach and established the infrastructure for collaborative work of LAMBDA teachers/trainers (see Figure 1, providers) with PhD students and other interested parties (see Figure 1,

Consumers). Activities are aligned also with the goals, strategies and recommendations defined by the European Big Data community in their work, see the *Strategic Research Innovation and Deployment Agenda for the AI, Data and Robotics* (SRIDA [5]).

Figure 1: LAMBDA Ecosystem

This paper describes the infrastructure that was established to reinforce organizational learning and capacity building at the Institute Mihajlo Pupin (PUPIN) and to facilitate teachers-trainees cooperation in the larger network of experts in the field of Big data, semantic technologies, Enterprise Knowledge Graphs (EKGs), Semantic Big Data Architectures (ARCH), and smart Big Data Analytics (BDA). The paper is structured as follows. Section 2 introduces the LAMBDA project and LAMBDA platform. Section 3 points to Big Data tools that were tested and evaluated in the LAMBDA framework. Section 4 provides an overview of the newly developed lectures on topics from the Big Data Analytics domain. Section 5 concludes the paper.

2. LAMBDA PLATFORM CUSTOMIZATION

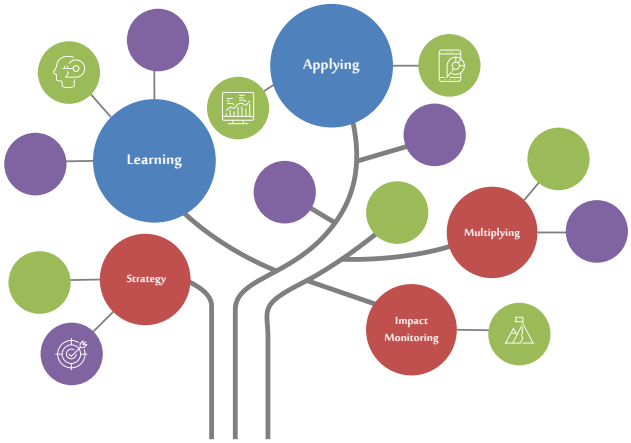

Figure 2: Main pillars of LAMBDA project

LAMBDA activities are structures around 3 main streams, as is presented in Figure 2:

- **Learning** – with the goal to establish and maintain a Knowledge repository as part of the LAMBDA Learning and Consulting Platform to facilitate spreading learning materials, as well as exchange of best practice;
- **Applying** – knowledge transfer and expertise exchange activities aimed at development of new prrof-of-concepts;
- **Multiplying** - raising awareness about future trends in Big Data, Emerging Tools and Technologies.

Hence, in order to support the transfer of institutional knowledge and expertise, but also to other relevant stakeholders in the region, the *LAMBDA Learning and Consulting Platform* was established using the Drupal content management system (CMS, https://www.drupal.org/). Since July 2018, it facilitates collaboration between consortium partners, e.g. joined paper and deliverable writing, information sharing, and stakeholders' data-base management. It is based on Drupal, the 3rd most popular Content Management System besides WordPress and Joomla. Some of the advantages of Drupal are related to the availability of different features / functionalities and customization, security and user experience i.e. Drupal is less resource-intensive and its pages typically load quicker and have faster response times than WordPress and Joomla. The architecture of Drupal CMS is modular and it consists of a few core modules and a plethora of plug-ins that can be switched on and off as required. In order to support the needs many different types of contents were defined, where each content type is described with a set of attributes, for instance, lecture is described with *category*, *contributor*, *format* (video, PPT, paper, book chapter). The platform provides access to the SlideWiki *open courseware* system, https://slidewiki.org/. The SlideWiki tool has the ability to import and export data from/into different data formats, thus SlideWiki users that use the SlideWiki presentation mode can merge and LAMBDA contents in their presentations.

3. OVERVIEW OF BIG DATA TECHNOLOGIES

Big Data refers to data sets which have large sizes and complex structures, while Big Data Analytics refers to the strategy of analysing large volumes of data that gathered from a wide variety of sources, including social networks, transaction records, videos, digital images and different kind of sensors. While more than 800,000 Petabyte (1 PB= 1015bytes) of data were stored in the year 2000, this volume will exceeds 175 zettabytes by 2025 as per International Data Corporation IDC expectations [6]. Some challenges related to the European ability to exploit the potential of Big Data are (1) fragmentation of the data ecosystem, due to different national policies, languages,

and sectors involved; (2) fragmentation of data research efforts and lack of effective exchange of results; and (3) shortage of highly skilled persons for data-related jobs. Therefore, the LAMBDA research was devoted to **demystifying Big Data topics**, including the latest achivement in the **field of Knowledge Graphs** [7, 8, 9].

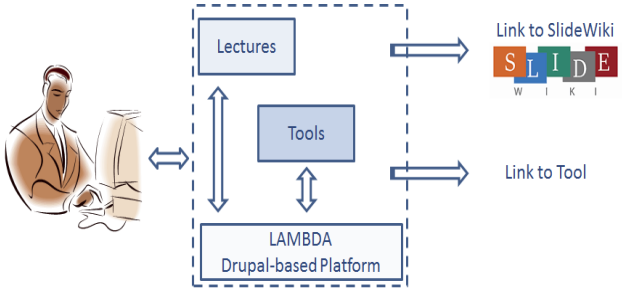

Figure 3: LAMBDA Platform - simplified illustration

Big Data Tools: One section of the LAMBDA Platform is devoted to tools and technologies. The tools have been categorized into twelve categories, see also Table 1 in the Annex: Cloud Marketplaces, Hadoop as a Web Service / Platform, Operational Database Management Systems, NoSQL/ Graph databases, Analytics Software / System / Platform, Data Analytics Languages, Optimization Library for Big Data, Library / API for Big Data, ML Library / API for Big Data, Visualization Software / System, and Distributed Messaging System.

Enterprise Knowledge Graphs (KGs) gained popularity in 2012 with the announcement of the Google Knowledge Graph as a representation of general world knowledge. KGs is a solution that allows to build a common understanding of heterogeneous, distributed data in organizations and value chains and thus to provide smart data for AI applications. However, many factors have prevented effective large-scale development and implementation of complex knowledge-based scenario because of inability to cope with the rising challenges coming from the Big Data applications; the rigidity of existing database management systems, inability to go beyond the standard requirements of query answering; and the lack of knowledge languages expressive enough to address real-world cases.

4. OVERVIEW OF LAMBDA LECTURES

One section of the LAMBDA Platform is devoted to open learning. Currently, the main providers of lectures are consortium partners and invited speakers at LAMBDA events relevant for LAMBDA are the University of Bonn (UBO), The University of Oxford (UOXF), the German National Library for Science and Technology (TIB) and the Institute Mihajlo Pupin:

- The *University of Bonn / Fraunhofer Institute* team works on the cutting edge technologies related to Big Data, Intelligent Analysis and Information Systems. The concerned team at the Smart Data Analytics (SDA) group is active in specializing applied research in intelligent data and knowledge analysis and teaching activities of the relevant topics. The *Fraunhofer* team is looking forward to applying the results of LAMBDA in different industrial domains.

- The *University of Oxford* team works on cutting edge technologies related to Big Data and analytics. The concerned team at the VADA ("Value Added Data Systems") group is active in research and teaching activities with regard to these topics. The *University of Oxford* team is looking forward to applying the results of LAMBDA in the financial domain, see for instance an example of adoption of LAMBDA lectures for vocational training, depicted in Figure 4.

- The *German National Library for Science and Technology* team is working on development of cutting edge technologies semantic data processing, knowledge engineering and information systems in different domains (including health).

- The *Institute Mihajlo Pupin* team is working on development of novel data analytics algorithms for different industrial domains, with main focus on the energy sector.

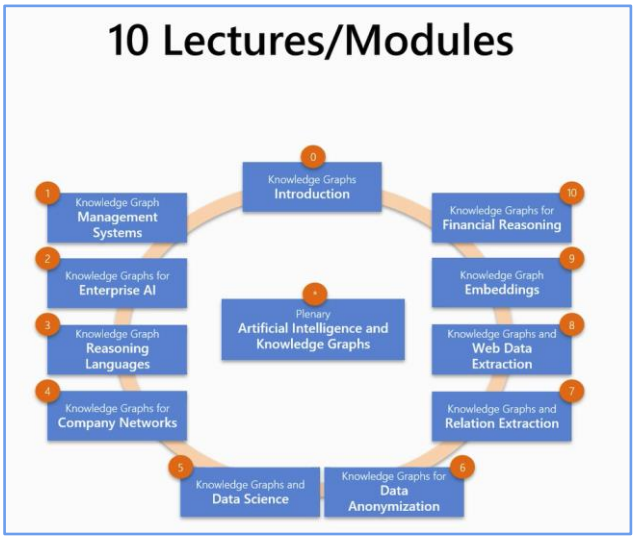

Figure 4: Adoption of LAMBDA lectures for finance domain (University of Oxford)

Since June 2020, the lectures have been categorized into eight modules, see also Table 2 in the Annex:

1. Artificial Intelligence (4 lectures),
2. Survey (3 lectures),
3. Foundations (3 lectures),
4. Enterprise Knowledge Graphs (4 lectures),
5. Semantic Big Data Architectures (7 lectures),

6. Big Data and Knowledge Graphs Tools (4 lecture),
7. Smart Data Analytics (5 lectures), and
8. Case Studies (5 lectures).

Key Performance Indicators (KPIs) for the self-evaluation of LAMBDA activities are given in Table 3 in the Appendix.

5. CONCLUSION

The potential behind the exploitation of data (Open, Linked and Big) to boost economies and growth has been in the focus of many EU initiatives, the most recent of which is the Digital Single Market, which highlights the need to make sense of Big Data, since this is considered to be a fertile ground for innovation in both technology and development. The EU funded project LAMBDA (Learning, Applying, Multiplying Big Data Analytics) addresses challenges related to Big Data Analytics and the semantics-based approach to processing data (Linked Data, Open Data, Big Data). The implemented activities in the LAMBDA framework (open education, research-industry collaboration) have strengthen the digital skills of professionals and improved the technologies and services of the involved stakeholders (PUPIN and other stakeholders from West Balkan), thus contributing to national and regional sustainable development. The LAMBDA researchers are currently involved in transfer of the knowledge and technologies to industry. To that aims, the LAMBDA Learning and Consulting Platform, https://project-lambda.org have been established that informes about the activities of the LAMBDA consortium and support the exchange of learning materials, tools, project results and best practice between the international leading organizations and research institutions and Industry from the West Balkan countries.

ACKNOWLEDGEMENT

The research presented in this paper is partly financed by the Ministry of Science and Technological Development of the Republic of Serbia and partly by the H2020 project "Learning, Applying, Multiplying Big Data Analytics" (GA No. 809965).

REFERENCES

[1] T. Allas et al., "Artificial intelligence is getting ready for business, but are businesses ready for AI?" in Driving impact at scale from automation and AI, Digital/McKinsey, February 2019.

[2] European Commission, "European Strategy for Data," February 2020, https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy.

[3] European Commission, "European Open Science Cloud," 2016, https://ec.europa.eu/digital-single-market/en/european-open-science-cloud,

[4] BDVA Association, "BDVA i-Spaces, 2018, https://www.bdva.eu/I-Spaces.

[5] BDVA Association, "Strategic Research, Innovation and Deployment Agenda (third release), September 2020.

A. Chi Zhou and B. He, "Big Data and Exascale Computing," in Sherif Sakr, Albert Y. Zomaya (eds) Encyclopedia of Big Data Technologies, Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-77525-8

[2] P. Zikopoulos and C. Eaton, "Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data," 2011

[3] A. Patrizio, "IDC: Expect 175 zettabytes of data worldwide by 2025", Network World, December 03, 2018, https://www.networkworld.com/article/3325397/idc-expect-175-zettabytes-of-data-worldwide-by-2025.html

[4] F. G. Filip and E. Herrera-Viedma, Big Data in the European Union, The Bridge Vol. 44, No. 4, Winter 2014, https://www.nae.edu/Publications/Bridge/128772/129172.aspx

[5] European Commission, Building a European data economy, https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy

[6] A. Patrizio, "IDC: Expect 175 zettabytes of data worldwide by 2025", Network World, December 03, 2018.

[7] V. Janev, D. Graux, H. Jabeen, E. Sallinger (Eds.) "Knowledge Graphs and Big Data Processing," Lecture Notes in Computer Science vol. 12072, pp. 1-208. Springer International Publishing. ISBN 978-3-030-53198-0. DOI: https://doi.org/10.1007/978-3-030-53199-7

[8] S. Auer, "Towards Knowledge Graph based Representation, Augmentation and Exploration of Scholarly Communications," Keynote at the 1st LAMBDA Big Data Analytics Summer School (18-20.06.2019, Belgrade, Serbia

[9] H. Paulheim, "Knowledge graph refinement: a survey of approaches and evaluation methods," Semantic Web Journal 8(3), 489–508, 2017.

APPENDIX: TABLES

Table 1: Categorization of tools

Category	Tools
Cloud Marketplaces	ALIBABA CLOUD; IBM CLOUD; GOOGLE CLOUD PLATFORM; ORACLE CLOUD MARKETPLACE; CISCO MARKETPLACE; MICROSOFT AZURE MARKETPLACE; AWS MARKETPLACE
Hadoop as a Web Service / Platform	HDINSIGHT; IBM INFOSPHERE BIGINSIGHTS; MAPR; CLOUDERA CDH; AMAZON EMR
Operational Database Management Systems	IBM (DB2); SAP (SAP HANA) ; MICROSOFT (SQL SERVER) ; ORACLE (DATABASE)
NoSQL/ Graph databases	HADOOP DISTRIBUTED FILE SYSTEM (HDFS) ; AMAZON NEPTUNE NEO4J; TIGERGRAPH; MAPR DATABASE; ONTOTEXT GRAPHDB; ALEGROGRAPH; VIRTUOSO; APPACHE JENA; MARKLOGIC JANUSGRAPH; ORIENTDB; MICROSOFT AZZURE COSMOS DB; APACHE HBASE; APACHE CASSANDRA; MONGODB
Stream Processing Engines	APACHE FLUME; APACHE APEX; AMAZON KINESIS STREAMS; APACHE FLINK; APACHE SAMZA; APACHE STORM; APACHE SPARK
Analytics Software / System / Platform	SAS ANALYTICS SOFTWARE & SOLUTIONS; MATLAB; H2O.AI; ACCORD FRAMEWORK; APACHE HADOOP; CLOUDERA DATA PLATFORM; VADALOG SYSTEM; MATLAB; SEMANTIC ANALYTICS STACK (SANSA)
Data Analytics Languages	SCALA; JULIA; SPARQL; SQL; R; PYTHON PACKAGE INDEX (PYPI) ; PYTHON
Optimization Library for Big Data	FACEBOOK AX; HYPEROPT; IBM ILOG CPLEX OPTIMIZATION LIBRARY
Library / API for Big Data	TENSORFLOW SERVING; MLLIB; BIGML; GOOGLE PREDICTION API; AZURE MACHINE LEARNING; AMAZON MACHINE LEARNING API; IBM WATSON PROGRAMMING WITH BIG DATA IN R
ML Library / API for Big Data	CAFFE.AI; APPACHE MXNET; XGBOOST; PYTORCH; KERAS; TENSORFLOW
Visualization Software / System	ORACLE VISUAL ANALYZER; MICROSOFT POWER BI; DATAWRAPPER; QLIKVIEW; CANVAS.JS; HIGHCHARTS; FUSION CHART; D3; TABLEAU; GOOGLE CHART
Distributed Messaging System	APACHE KAFKA

Table 2: Overview of lectures

Category	Tools
Survey (3)	Survey on Big Data Tools; Overview and Comparison of Machine Learning Algorithms; Survey on Big Data Applications
Artificial Intelligence (4)	Data for AI: Foresight; AI and Knowledge Graphs; Conversational AI; The Revolution of AI
Foundations (3)	Big Data Ecosystem; Introduction to Knowledge Graphs; Big Data Outlook, Tools, and Architectures
Enterprise Knowledge Graphs (5)	What is Knowledge Graph? Introduction to Knowledge Graphs; Creation of Knowledge Graphs; Extraction for Knowledge Graphs; Swift Logic for Big Data and Knowledge Graphs
Semantic Big Data Architectures (5)	Reasoning in Knowledge Graphs; Introduction to Big Data Architecture;

	Big Data Solutions in Practical Use-cases; Distributed Big Data Frameworks; Data Lakes and Federated Query Processing
Big Data and Knowledge Graphs Tools (4)	Context-Based Entity Matching for Big Data; Vadalog System; Data Science with Spark and Hadoop; Spark using Scala
Smart Data Analytics (5)	Distributed Big Data Libraries; Distributed Semantic Analytics I; Distributed Semantic Analytics II; SANSA - Scalable Semantic Analytics Stack; Scalable Knowledge Graph Processing using SANSA
Case Studies (9)	Semantic Information Infrastructures from Business Information Delivery to Water Management; Soft computing for Transparent synthesis of Geo Big Data; Chronorobotics - Spatio-temporal models for social and service robots; IntelliSys: Intelligent System for Road Safety; Reasoning on Financial Knowledge Graphs: The Case of Company Networks; Embedding-based Recommendations on Scholarly Knowledge Graphs; Open and Big Data – Utilization Perspective; Data Analytics for Energy Sector; Predictive Analytics in Renewable Energy Systems

Table 3: Key Performance Indicators

Category	Success Indicator
STRATEGIES / RECOMMENDATIONS	number of strategic documents
BIG DATA ANALYTICS SCHOOL	number of events
	number of lectures (books, published, e.g. via ceur-ws.org)
	number of trained teachers / students
LAMBDA-NETWORK OF EXPERTS	number of organizations
	number of experts
CAPACITY BUILDING THROUGH THE LAMBDA LEARNING AND CONSULTING PLATFORM	number of tools integrated for experimentation
	number of visits / exchanges between partners and collaborators.
	number of university / organizations that have adopted the developed open lectures
DISSEMINATION AND OUTREACH (OTHER THAN BDA SCHOOL)	number of workshops, seminars and networking events organized
	number of brainstorming sessions on key society challenges
	number of joint scientific papers