A Big Data Learning Platform for the West Balkans and Beyond

Damien Graux  
ADAPT Centre, Trinity College Dublin, Ireland  
damien.graux@adaptcentre.ie

Hajira Jabeen  
University of Cologne, Germany  
hajira.jabeen@uni-koeln.de

Valentina Janev  
The Mihajlo Pupin Institute, University of Belgrade, Serbia  
valentina.janev@pupin.rs

Emanuel Sallinger  
TU Wien, Austria and University of Oxford, UK  
emanuel.sallinger@cs.ox.ac.uk

ABSTRACT
As the number of Big Data related methods, tools, frameworks and solutions is growing, there is a need to classify and make available the knowledge related to this domain. This is especially useful for countries which, so far, have suffered from some lack of infrastructure in Big Data. In this article, we describe the deployment of an open online platform gathering lectures and resources, from multiple partnered European institutions, tailored for West Balkan students through focuses on local Big Data challenges.

CCS CONCEPTS
- Information systems → Data management systems; → Applied computing → Education.

KEYWORDS
Teaching Big Data Analytics, West Balkans, European collaboration

1 INTRODUCTION
Big Data refers to data sets which have large sizes and complex structures. The data size can range from dozens of terabytes to a few zettabytes and is still growing [7]. Big Data Analytics, hence, refers to the strategy of analyzing 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. In an attempt to support the European data economy policy [1], our consortium proposed a training approach [2] and established the infrastructure for collaborative work of teachers/trainees with PhD students and other interested parties such as industries.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ITiCSE'21, 2021, Germany
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06 ... $15.00
https://doi.org/10.1145/1122445.1122456

1 https://www.linkedin.com/groups/12129621/
2 Downloaded more than 40 000 times as of January 2021.
3 https://www.drupal.org/
handle Big Data, we proposed a characterization of the landscapes in the topics related to Big Data. Indeed, we propose to categorise the Big Data landscape as follows: Cloud Marketplaces; Hadoop as a Web Service/Platform; Operational Database Management Systems; NoSQL/Graph databases; Stream Processing Engines; 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; Distributed Messaging System. Overall, more than 80 tools have been identified and classified so that the students can be guided by their use-case when they need to select a tool for a specific application.

2.4 An OpenCourseWare Integration

The platform provides access to the SlideWiki OpenCourseWare system.6 In a nutshell, SlideWiki is an open, web-based OpenCourseWare authoring platform that aims to provide an open and accessible platform to create and share qualitative, rich and engaging educational content following the 5R principles of OERs [6]. The platform allows educators to create, edit, translate and reuse HTML slide presentations complemented with comments, links to sources and supporting materials as well as questions to help learners.

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 the content we are sharing directly into their presentations. Moreover, SlideWiki allows to create self-assessment exam sheets; therefore, teachers are able to set up exams where users can test their knowledge about a lecture.

3 CONCLUSIONS

In order to foster knowledge about Big Data analytics in the West Balkan area, we set up various actions, from staff-exchange to summer schools. Among these actions, we also deployed an online and open platform gathering all the material presented to students plus additional analyses. So far, our platform counts more than 150 registered users and shares 40+ lectures in different formats. Moreover, the initiative of opening such a platform allows users to learn at anytime virtually (which proved to be very useful during the COVID-19 outbreak) and from anywhere. The latter thereby broadens the possible audience of our lectures which can now be taken outside the West Balkan region.

REFERENCES

[8] https://slidewiki.org/