

# Overview of SIMBig 2015: 2nd Annual International Symposium on Information Management and Big Data

**Juan Antonio Lossio-Ventura**  
LIRMM, University of Montpellier  
Montpellier, France  
juan.lossio@lirmm.fr

**Hugo Alatrística-Salas**  
Pontificia Universidad Católica del Perú  
Lima, Peru  
halatrística@pucp.pe

## Abstract

Big Data is a popular term used to describe the exponential growth and availability of both structured and unstructured data. The aim of the symposium is to present the analysis methods for managing large volumes of data through techniques of artificial intelligence and data mining. Bringing together main national and international actors in the decision-making field to state in new technologies dedicated to handle large amount of information.

## 1 Introduction

Big Data is a popular term used to describe the exponential growth and availability of both structured and unstructured data. This has taken place over the last 20 years. For instance, social networks such as Facebook, Twitter and LinkedIn generate masses of data, which is available to be accessed by other applications. Several domains, including biomedicine, life sciences and scientific research, have been affected by Big Data<sup>1</sup>. Therefore there is a need to understand and exploit this data. This process can be carried out thanks to “Big Data Analytics” methodologies, which are based on Data Mining, Natural Language Processing, etc. That allows us to gain new insight through data-driven research (Madden, 2012; Embley and Liddle, 2013). A major problem hampering Big Data Analytics development is the need to process several types of data, such as structured, numeric and unstructured data (e.g. video, audio, text, image, etc)<sup>2</sup>.

Therefore, the second edition of the Annual International Symposium on Information Management and Big Data - SIMBig 2015<sup>3</sup>, aims to present the analysis methods for managing large volumes of data through techniques of artificial intelligence and data mining. Counting with main national and

international actors in the decision-making field to state in new technologies dedicated to handle large amount of information.

Our first edition, SIMBig 2014<sup>4</sup> took place in Cuzco Peru too in September 2015. SIMBig 2014 has been indexed on DBLP<sup>5</sup> (Lossio-Ventura and Alatrística-Salas, 2014) and on CEUR Workshop Proceedings<sup>6</sup>.

### 1.1 Keynote Speakers

SIMBig 2015 second edition has welcomed five keynote speakers experts in Big Data, Data Mining, Natural Language Processing (NLP), and Social Networks:

- PhD. Pr. **Albert Bifet**, from HUAWEI Noah’s Ark Lab, China;
- PhD. Pr. **Diana Inkpen**, from University of Ottawa, Canada;
- PhD. Pr. **Pascal Poncelet**, from LIRMM Laboratory and University of Montpellier, France;
- PhD. Pr. **Mathieu Roche**, from Cirad and TETIS Laboratory, France;
- PhD. Pr. **Osmar R. Zaiane**, from University of Alberta, Canada.

### 1.2 Scope and Topics

To share the new analysis methods for managing large volumes of data, we encouraged participation from researchers in all fields related to Big Data, Data Mining, and Natural Language Processing, but also Multilingual Text Processing, Biomedical NLP. Topics of interest of SIMBig 2015 included but were not limited to:

- Big Data
- Data Mining
- Natural Language Processing

<sup>1</sup>By 2015 the average of data annually generated in hospitals is 665TB: <http://ihealthtran.com/wordpress/2013/03/infographic-friday-the-body-as-a-source-of-big-data/>.

<sup>2</sup>Today, 80% of data is unstructured such as images, video, and notes

<sup>3</sup><http://simbig.org/SIMBig2015/>

<sup>4</sup><https://www.lirmm.fr/simbig2014/>

<sup>5</sup><http://dblp2.uni-trier.de/db/conf/simbig/simbig2014>

<sup>6</sup><http://ceur-ws.org/Vol-1318/index.html>

- Bio NLP
- Text Mining
- Information Retrieval
- Machine Learning
- Semantic Web
- Ontologies
- Web Mining
- Knowledge Representation and Linked Open Data
- Social Networks, Social Web, and Web Science
- Information visualization
- OLAP, Data Warehousing
- Business Intelligence
- Spatiotemporal Data
- Health Care
- Agent-based Systems
- Reasoning and Logic
- Constraints, Satisfiability, and Search

## 2 Latin American and Peruvian Academic Goals of the Symposium

The academic goals of the symposium are varied, among which we can list the following:

- Meet Latin American and foreign researchers, teachers, and students belonging to several domains of computer sciences, specially related to Big Data.
- Promote the production of scientific articles, which will be evaluated by the international scientific community, in order to receive a feedback from experts.
- Foster partnerships between Latin American universities, local universities and European universities.
- Promote the creation of alliances between Peruvian universities, enabling decentralization of education.
- Motivate students to learn more about computer sciences research to solve problems related to the management of information and Big Data.
- Promote the research in Peruvian universities, mainly those belonging to the local organizing committee.
- Create connections, forming networks of partnerships between companies and universities.
- Promote the local and international tourism, in order to show to the participants the architecture, gastronomy and local cultural heritage.

## 3 Track on Web and Text Intelligence (WTI 2015)

Web and text intelligence are related areas that have been used to improve human computer interaction both in general and in particular to explore and analyze information that is available on the Internet. With the advent of social networks and the emergence of services such as Facebook, Twitter, and others, research in these areas has been greatly enhanced. In recent years, shared knowledge and experiences have established new and different types of personal and communal relationships which have been leveraged by social networks scientists to produce new insights. In addition there has been a huge increase in community activities on social networks.

The Web and Text Intelligence (WTI) track of SIMBig 2015 have provided a forum that brought together researchers and practitioners for exploring technologies, issues, experiences and applications that help us to understand the Web and to build automatic tools to better exploit this complex environment. The WTI track has fostered collaborations, exchange of ideas and experiences among people working in a variety of highly cross-disciplinary research fields such as computer science, linguistics, statistics, sociology, economics, and business.

The WTI track is a follow up of the 4th International Workshop on Web and Text Intelligence<sup>7</sup>, which took place in Curitiba, Brazil, October 2012, as a workshop of BRACIS 2012; the 3rd International Workshop on Web and Text Intelligence<sup>8</sup>, which took place in São Bernardo, Brazil, October 2010, as a workshop of SBIA10; the 2nd International Workshop on Web and Text Intelligence<sup>9</sup>, which took place in São Carlos, Brazil, September 2009, as a workshop of STIL09; the 1st Web and Network Intelligence<sup>10</sup>, which took place in Aveiro, Portugal, October 2009, as a thematic track of EPIA09; and the 1st International Workshop on Web and Text Intelligence, which took place in

<sup>7</sup><http://www.labic.icmc.usp.br/wti2012/>

<sup>8</sup><http://www.labic.icmc.usp.br/wti2010/>

<sup>9</sup><http://www.labic.icmc.usp.br/wti2009/>

<sup>10</sup><http://epia2009.web.ua.pt/wni/>

Salvador, Brazil, October 2008, as a workshop of SBIA08.

### 3.1 Scope and Topics

The topics of WTI include, but are not limited to:

- Web and Text Mining
- Link Mining
- Web usability
- Web automation and adaptation
- Graph and complex network mining
- Communities analysis in social networks
- Relationships analysis in social networks
- Applications of social networks and social media
- Data modeling for social networks and social media
- Location-based social networks analysis
- Big data issues in social network and media analysis
- Modeling of user behavior and interactions
- Temporal analysis of social networks and social media
- Pattern analysis in social networks and social media
- Privacy and security in social networks
- Propagation and diffusion of information in social network
- Social information applied to recommender systems
- Search and Web Mining
- Multimedia Web Mining
- Visualization of social information

## 4 Sponsors

We want to thank our wonderful sponsors! We extend our sincere appreciation to our sponsors, without whom our symposium would not be possible. They showed their commitment to making our research communities more active. We invite you to support these community-minded organizations.

### 4.1 Organizing Institutions

- Université de Montpellier, France<sup>11</sup>
- Laboratoire de Informatique, Robotique et Microélectronique de Montpellier, France<sup>12</sup>
- Universidad Andina del Cusco, Perú<sup>13</sup>

<sup>11</sup><http://www.umontpellier.fr/>

<sup>12</sup><http://www.lirmm.fr/>

<sup>13</sup><http://www.uandina.edu.pe/>

### 4.2 Collaborating Institutions

- iMedia<sup>14</sup>
- Bioincuba<sup>15</sup>
- TechnoPark<sup>16</sup>
- Grupo de Reconocimiento de Patrones e Inteligencia Artificial Aplicada, PUCP, Perú<sup>17</sup>
- Universidad Nacional Mayor de San Marcos, Perú<sup>18</sup>
- Escuela de Post-grado de la Pontificia Universidad Católica del Perú<sup>19</sup>

### 4.3 WTI Organizing Institutions

- Instituto de Ciências Matemáticas e de Computação, USP, Brasil<sup>20</sup>
- Instituto Politécnico de Tomar, Portugal<sup>21</sup>
- Laboratório de Inteligência Computacional, ICMC, USP, Brasil<sup>22</sup>
- Laboratório de Inteligência Artificial e Apoio à Decisão, INESC TEC, Portugal<sup>23</sup>
- Machine Learning Lab (MaLL), UFSCar, Brasil<sup>24</sup>
- Universidade Federal de São Carlos, Brasil<sup>25</sup>

## References

David W Embley and Stephen W Liddle. 2013. Big data—conceptual modeling to the rescue. In *Conceptual Modeling*, ER’13, pages 1–8. LNCS, Springer.

Juan Antonio Lossio-Ventura and Hugo Alatrística-Salas, editors. 2014. *Proceedings of the 1st Symposium on Information Management and Big Data - SIMBig 2014, Cusco, Peru, September 8-10, 2014*, volume 1318 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Sam Madden. 2012. From databases to big data. volume 16, pages 4–6. IEEE Educational Activities Department, Piscataway, NJ, USA, may.

<sup>14</sup><http://www.imedia.pe/>

<sup>15</sup><http://www.bioincuba.com/>

<sup>16</sup><http://technoparkidi.org/>

<sup>17</sup><http://inform.pucp.edu.pe/~grpiaa/>

<sup>18</sup><http://www.unmsm.edu.pe/>

<sup>19</sup><http://posgrado.pucp.edu.pe/la-escuela/presentacion/>

<sup>20</sup><http://www.icmc.usp.br/Portal/>

<sup>21</sup><http://portal2.ipt.pt/>

<sup>22</sup><http://labic.icmc.usp.br/>

<sup>23</sup><http://www.inesctec.pt/liaad>

<sup>24</sup><http://ppgcc.dc.ufscar.br/pesquisa/laboratorios-e-grupos-de-pesquisa>

<sup>25</sup><http://www2.ufscar.br/home/index.php>