Preface

The aim of the SIMBig symposium is to present the analysis of methods for extracting knowledge from large volumes of data through techniques of data science and artificial intelligence. This books comprises extended versions of the best papers presented at SIMBig 2015 and SIMBig 2016.

Big data is a popular term used to describe the exponential growth and availability of data, which could be structured and unstructured. Data science is a field seeking to extract knowledge or insights from large volumes of heterogeneous data (e.g., video, audio, text, image). Data science is a continuation of other fields such as data analysis, statistics, machine learning, and data mining similar to knowledge discovery in databases (KDD).

Big data has taken place over the past 20 years. For instance, social networks such as Facebook, Twitter, and LinkedIn generate masses of data, which are available to be accessed by other applications. Several domains, including biomedicine, life sciences, and scientific research, have been affected by big data[1]. Therefore, there is a need to understand and exploit these data. This process can be carried out thanks to "data science", which is based on methodologies of data mining, natural language processing, Semantic Web, statistics, etc. That allows us to gain new insight through data-driven research [1, 4]. 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.)[2].

Our Annual International Symposium on Information Management and Big Data, seeks to present the new methods of data science and related fields for analyzing and managing large volumes of data. The symposium attracts many of the main national and international players in the decision-making field who are involved in new technologies dedicated to handling large amounts of information.

The third edition, SIMBig 2016[3], was held in Cusco, Peru, during September 1–3, 2016. SIMBig 2016 has been indexed in DBLP[4] [3] and in the CEUR Workshop Proceedings[5]. The second edition, SIMBig 2015[6], was also held in Cusco, Peru, during September 2–4, 2015. SIMBig 2015 has been indexed in DBLP[7] [2] and in the CEUR Workshop Proceedings[8].

---

[1] By 2015 the average amount of data generated annually in hospitals is 665TB: https://datafloq.com/read/body-source-big-data-infographic/413.

[2] Today, 80% of data are unstructured such as images, video, and notes.

[3] http://simbig.org/SIMBig2016/.

[4] http://dblp2.uni-trier.de/db/conf/simbig/simbig2016.html.

[5] http://ceur-ws.org/Vol-1743/.

[6] http://simbig.org/SIMBig2015/.

[7] http://dblp2.uni-trier.de/db/conf/simbig/simbig2015.html.

[8] http://ceur-ws.org/Vol-1478/.

For this special proceedings volume, we accepted 11 long papers. These papers were selected from the SIMBig 2015 and SIMBig 2016 editions. We selected the four best papers from the 2015 edition, which had 32 submissions. Likewise, we selected the seven best papers of SIMBig 2016, from 42 submissions. Therefore, the general acceptance rate of this special proceedings volume was $11/(42 + 32) = 0.15$.

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 science, data mining, natural language processing, and the Semantic Web, but also multilingual text processing as well as biomedical NLP.

Topics of interest of SIMBig included: data science, big data, data mining, natural language processing, 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.

February 2017   Juan Antonio Lossio-Ventura
                                                                       Hugo Alatrista-Salas

References

1. David W Embley and Stephen W Liddle, *Big data—conceptual modeling to the rescue*, Conceptual Modeling, ER'13, LNCS, Springer, 2013, pp. 1–8.
2. Juan Antonio Lossio-Ventura and Hugo Alatrista-Salas (eds.), *Proceedings of the 2nd annual international symposium on information management and big data - simbig 2015, cusco, Peru, September 2–4, 2015*, CEUR Workshop Proceedings, vol. 1478, CEUR-WS.org, 2015.
3. Juan Antonio Lossio-Ventura and Hugo Alatrista-Salas (eds.), *Proceedings of the 3rd annual international symposium on information management and big data - simbig 2016, cusco, Peru, September 1–3, 2016*, CEUR Workshop Proceedings, vol. 1743, CEUR-WS.org, 2016.
4. Sam Madden, *From databases to big data*, vol. 16, IEEE Educational Activities Department, Piscataway, NJ, USA, May 2012, pp. 4–6.

Information Management and Big Data
Second Annual International Symposium, SIMBig 2015,
Cusco, Peru, September 2-4, 2015, and Third Annual
International Symposium, SIMBig 2016, Cusco, Peru,
September 1-3, 2016, Revised Selected Papers
Lossio-Ventura, J.A.; Alatrista-Salas, H. (Eds.)
2017, XI, 147 p. 46 illus., Softcover
ISBN: 978-3-319-55208-8