ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/323059235

Data-driven decision making

Conference Paper · December 2017

DOI: 10.1109/ICTUS.2017.8285973

CITATIONS
19

READS
13,369

1 author:

Mario José Diván
Intel
107 PUBLICATIONS **324** CITATIONS

SEE PROFILE

All content following this page was uploaded by Mario José Diván on 09 July 2020.

The user has requested enhancement of the downloaded file.

Data-Driven Decision Making

Mario José Diván[#*]

[#]*Economics and Law School, University of La Pampa*
Coronel Gil 353, 1st Floor, Santa Rosa (La Pampa, Argentina)
[1]mjdivan@[eco|ing].unlpam.edu.ar

[*]Engineering School, University of La Pampa
Street 9 and 110, General Pico (La Pampa, Argentina)

Abstract— **Along the decision-making process we depend of assumptions, premises, the context and this is guided through the aim associated with the decision itself. The context and assumptions represents external aspects out of the control of any decision maker, but the premises and the knowledge of the company depends of our data. A common conceptual mistake is associated with the confusion related to data and information when indeed they are very different concepts. It is that to say, we can gather data from different heterogeneous data sources but nothing warranty us that the data are consistent, comparable and traceable. In this work we talk about the importance of the measurement and evaluation in relation with the data-driven decision making. Followed, we present the implications of an intuition-driven decision making and their possible social impact. Finally, An application case related with the monitoring of business process in the Autarchic Institute of Housing (La Pampa, Argentina) is shown for describing the application of the concepts related to data-driven decision making.**

Keywords— Data-Driven Decision Making, Measurement, Evaluation, Social Impact

I. INTRODUCTION

Along the decision-making process we depend of assumptions, premises, the context and this is guided through the aim associated with the decision itself [1]. The context and assumptions represents external aspects out of the control of any decision maker, but the premises and the knowledge of the company depends of our data because they are part of our organization as system [2]. A common conceptual mistake is associated with the confusion related to data and information when indeed they are very different concepts. It is that to say, we can gather data from different heterogeneous data sources but nothing warranty us that the data are consistent, comparable and traceable. In this sense, if we need make a decision, we need to know the entity under analysis and their associated information at the precise instant. It is important because from the point of view associated with the general system theory [1]- [2], it is necessary identifies the system for identifying their boundaries, context, subsystems, feedback, input and outputs. Once the system was identified, we can go on the quantification related to each associated characteristic for knowing it in detail.

Thus and for knowing the entity under analysis, we need to measure it for quantifying their associated characteristics and from there; we define the indicators for interpreting each metric's value. In this way, the Measurement and Evaluation (M&E) process can be supported by a conceptual framework with an underlying ontology. The M&E framework allows defining the necessary concepts for carry on a measurement process in consistent and repeatable way [3]- [4]- [5].

Even when is important that a measurement process give us consistent, comparable and traceable results, but also it is very important its automatization. In the now economy, the operations happen in real time and for that reason we need seriously consider the online monitoring for detecting and preventing different situations on the fly. In this sense, the role of the measurement and evaluation frameworks is a key asset, because they allow structure and automate the measurement process in consistent way [6].

Once than it is possible to warranty that the measures are comparable, consistent and traceable, the decision-making process will be naturally based on their history (the measures along the time). In this aspect, the Organizational Memory take a particular importance, because allows store the organizational experience and knowledge for future recommendations (i.e. as foundation of assumptions, premises, among others). The Organizational Memory is fed continuously by the measures and their associated experiences, and it constitutes the base for the feedback in the decision-making process [7].

However, the Organizational Memory is a model and for that reason, it is possible that there not exist recommendations (or experiences) for a new situation (i.e. natural disaster). It is very important for remembering because in cases associated with measurement and evaluation processes on infrastructure in the context of smart cities, it is possible get partial data because it is highly possible that there not exist previous records. The last is the case of the city of Santa Rosa (La Pampa, Argentina) where even with previous experience about the level of raining, nothing could be done when in a week the city received water in an equivalent volume to one-year [8].

In this invited talk we talk about the influence of the data and information along the decision-making process. Also, we focus the measurement and evaluation process as key asset associated for knowing the entities under analysis (e.g. a business process, a person, a system, etc.), their contexts, and the way in which the process could be automatized. We highlight the role associated with the Organizational Memory as knowledge base for recommendations.

This article is organized as follows. Section 2 synthesizes the importance associated with the measurement and evaluation as engine of the data-driven decision making. Section 3 introduces the social impact of the data-driven

)78-1-5386-0514-1/17/$31.00 ©2017 IEEE

decision making considering the smart cities. Section 4 presents an application case related to the Autarchic Institute of Housing (La Pampa, Argentina). Section 5, shows some related works and finally the conclusions and future works are detailed.

II. The importance of the Measurement and Evaluation

For start a discussion it is interesting to question the concepts and their applications. That is to say, what's happen if we do not measure? Why we need to measure? What are the benefits? The common sense in the engineering, say us that we need characterize a concept or an object for knowing their physical and abstract characteristics. Once we know each characteristic, it is useful to quantify each one for studying the behaviour en different situations. Thus and from the study of each situation, we can basically identify a normal and abnormal situation; which is useful for detecting and preventing not wanted results. In this sense, the avoidable situations and the optimization of resources give us an interesting social and economic point of view as positive argument of the measurement. In each case, the concerns associated with the quality of the information keep being an active branch of researching [9].

The measurement allows quantifying the characteristics of an entity under analysis, such as a system, a component, etc. However, we need put special attention of the concepts related with the measurement process for warranting the homogeneity [10]. That is to say, the measurement is useful if and only if the measures are consistent and comparable and the measurement process is repeatable. For that reason, we need have the same interpretation in relation to the concepts of measures, metrics, indicators, among others. In this aspect, the measurement and evaluation frameworks take special sense, because allow us make an agreement about the concepts that we want use along the measurement process and speak in the same language avoiding misunderstandings.

For example, if we want to monitor an organization as system, we could use the Balanced Scorecard perspective of Kaplan & Norton [11]- [12]; the Goal-Question Metric approach [13]; the C-INCAMI (acronym of Context-Information Need, Concept Model, Attribute, Metric and Indicator) framework [4]; among others. Each approach could have weakness and strengths, but it depends of the situation in which we want apply them [14]. Independently of chosen approach, we need keep consistency along the time in relation to the way in which we measure for warranting the consistency and comparability of concepts and measures.

In terms of Data Quality based on the ISO 25.012 [15], there are characteristics related exclusively with the data itself, which implies that the characteristics depend exclusively of the data, such is the case of the accuracy, completeness, consistency, credibility and currentness. However, there are other characteristics which depend of the data and the system at the same time, such as the accessibility, compliance, confidentiality, efficiency, precision, among others. This is important to highlight because the data is a part of the system, and the data quality is affected by the data itself but also by the system which process them.

The data-driven decision making could be defined as "the practice of basing decisions on the analysis of the data rather than purely on intuition" [16]. In this sense, we can quickly note that if the decision-making is based on the data, a poor data quality will directly affect the decision-making process. For this reason, the monitoring on each stage of the data life cycle is critic. That is to say, we need to use practices in the organization for monitoring the data acquisition, data processing, data analysis, data preservation and data reuse, use or deletion [17]. In this sense, there are interesting proposals related to different perspectives associated with the needed practices for keeping the data quality, such as the data maturity model from CMMI Institute [18], or CALDEA model based on maturity models [19].

Coming back to the beginning of the section, Why is useful the measurement? It is useful for knowing the entity under analysis. The measurement define the gathering process and it is directly associated with the data acquisition in the data life cycle [20]. Consequently, if we put attention at the moment in which the data is gathered, we have serious possibility of to improve the other stages of the life cycle. In other words, if we could to decrement the possibility of error of the data at the source, then we could decrement the effect of the error propagation in the others stages associated with the life cycle. Finally, this would allow us to decrement the risks associated with data quality (e.g. consistency, etc.) when we make decisions based on the data.

However, the measurement refers to the way in we obtain the measures but not refer about how interpret them [21]. For example, in terms of C-INCAMI the evaluation suppose the formalization of the organizational knowledge through the decision criteria embedded in indicators [10]. In this way, each indicator has the enough concepts for interpreting each value of the associated metric and to arrive to a conclusion using the organizational knowledge.

Thus, the order in which we can define the measurement and evaluation strategy is important, and for that reason we can to take advantage of approaching such as GOCAME (Goal-Oriented Context-Aware Measurement and Evaluation) [4] and SiQinU (Strategy for Understanding and Improving Quality in Use) [6].

Fig. 1. Overflow of the Santo Tomas Lagoon on the city of Santa Rosa (La Pampa, Argentina). Photography by B.Dillon, “La Arena” Daily [22]

In terms of data-driven decision making the organization makes different decisions based on their data [16]. In this sense, the history associated with the data is a key asset for supporting the decision-making. For that reason the organizational memories could be addressed for modelling the organizational experiences from the historical measures and evaluations. Moreover, a case-based reasoning could be deployed from the organizational memory for supporting the recommendation in the decision-making process [7].

III. The Social Impact of the Data-Driven Decision Making

Smart city refers to cities which integrates and monitors the critical infrastructure using smart computing to deliver core services to public [17]. However, the idea behind of smart cities is not only related with a technological factor, but also is associated with aspects such as governance, economy, among others.

Each decision along the different services or infrastructure in a city should be sustained on their experiences. In this way, using the experiences and data from the city, we are able to typify different situations of normality and consequently, we can detect the situations out of normality. This is a key aspect for monitoring infrastructure and services in a city, and the final idea is related to prevent risk situations, and in the worst case, their detection in real time.

The measurement and evaluation process is critical for knowing the current state of each service or infrastructure in the city. That is to say, if we do not measure we will not know the state of each element. This is critical for orienting the data-decision making, because each decision should be based on the current situation of different elements along the city.

What if we do not have the data? It is highly possible that the decisions are oriented by the intuition since we do not have facts or records for supporting the alternatives of decision [1].

The problem associated with the intuition in the decision-making is that is subjective and we do not have records or previous experiences for justifying the course of actions [17].

When the authorities guide the decision-making in base of the intuition, the social impact could be a catastrophe. For example, the city of Santa Rosa is located in the province of La Pampa (Argentina). The city use a pluviometer as reference of the level of falling water in a determined raining. This allows gather data but not monitoring in real-time the situation related with the falling water along the city or the volume of water circulating along the sewers. During March of 2017 and only in a week, the city naturally received a volume of water equivalent to one-year of raining. As you can see in figure 1 [22], the consequences associated with the intuition in the decision-making was evident in the city of Santa Rosa; a picture say more than one thousand of words. At the date, the north zone and the capital of the province of

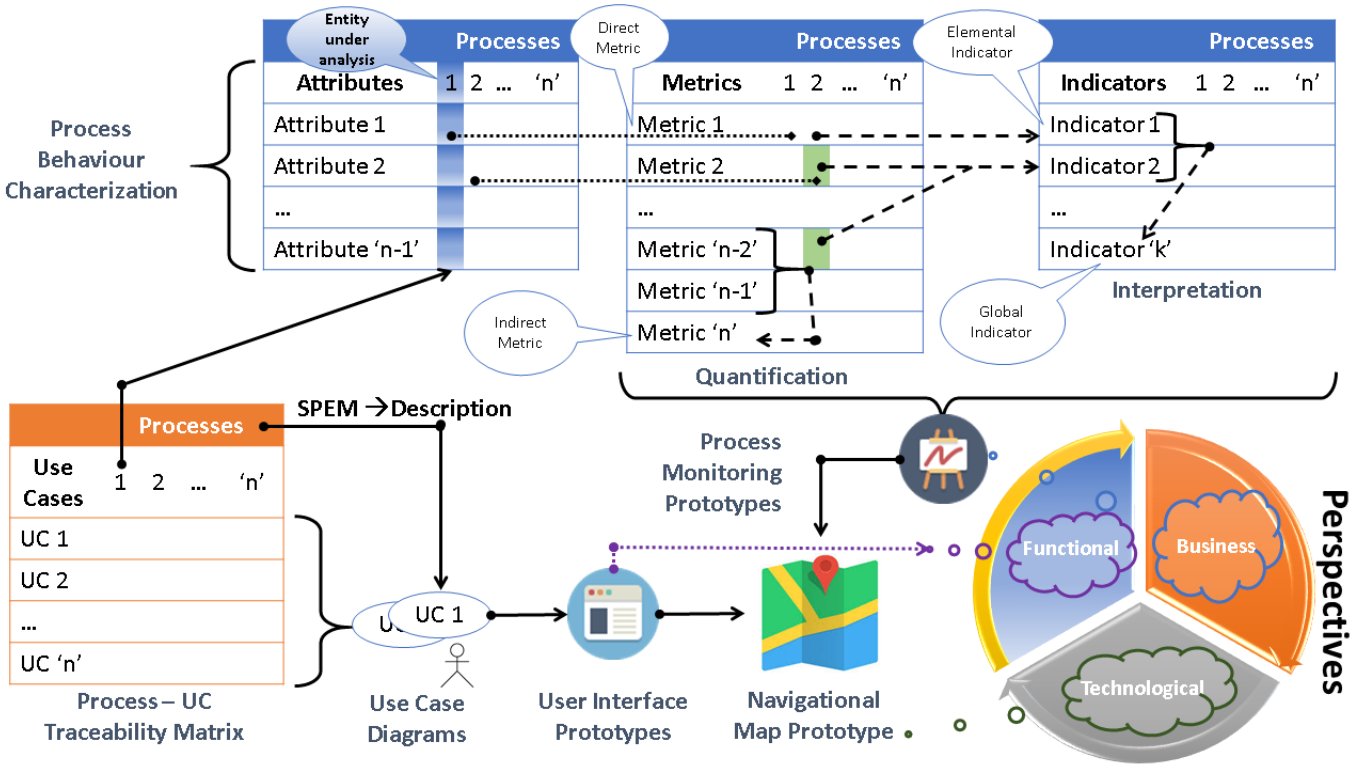

Fig. 2. Conceptual Perspective associated with the Business Process Monitoring. A Data-Driven Decision Making

La Pampa keep suffering the consequences of the inundations. Even when the natural disasters are not easy to prevent, it is possible to monitor the infrastructure and services along the city for planning the works with the due anticipation.

In the other extreme, the monitoring of the infrastructures and the services could at least improve the quality of life; anticipate disasters or even save the lives of the citizens.

IV. An Application Case: The Autarchic Institute of Housing of La Pampa

The Autarchic Institute of Housing of La Pampa (AIHLP) is the public organism responsible for the building of houses along all the provincial territory. Even when the organism receives funds from the provincial and national government, it administrates the funds in autarchic way. The future owners of each house are associated with families without the enough resources for accessing to the formal system carry on the banks. In this way, the associated demand is continuous and the offer is always lesser than the demand.

Given the high demand, the AIHLP has a specific normative for assigning the houses considering the particular situation of each family. In this way, the authorities wanted make decisions on the base of previous experience for learning from past situations, optimize the resources and to avoid the repetition of errors..

In 2014, we start a business process reengineering to be transparent the behavior of each process, using SPEM (acronym of Software Process Engineering Meta-Model) [23] as modeling language. We use EIbPREME (Integrated Strategy based on Processes, Requirements, Measurements and Evaluations), a process-oriented strategy, which the aim is related to monitoring the processes as entity under analysis using C-INCAMI as Measurement and Evaluation framework [24].

Once we managed to validate each process with our stakeholders, we started the definition of the measurement and evaluation project using the GOCAME strategy [6].

In this sense, our entity under analysis was each modeled process. From there, in coordination with the authorities, we define the mode in which they wanted characterize each business process and their associated point of view. The figure 2 synthesizes the idea from the identification of process attributes to the implementation of the three perspectives of monitoring (technological, functional and business).

As you can see in figure 2, each characteristic or attribute descriptive of a process was associated with a metric for its quantification. Followed, in collaboration with the authorities, an indicator was defined for interpreting each associated metric in each process. In this point, the business knowledge of the authorities was essential for incorporating the decision criteria inside the indicator definition.

From the metrics and indicators, we managed to prototype the visual scorecard for monitoring the processes. In this way, the business perspective concluded with the implementation of the web-enabled and multi-device command board.

In parallel, each task and activity in the process allow us to derive the model of use cases. From the uses cases the user interfaces were prototyped, and once they were validated by the final user, they were implemented.

Both the user interface and the command board are naturally linked because the processes incorporate the logical

view associated with the data gathering, data processing and data reporting.

Additionally to the business and logical view, the authorities incorporated the possibility to make the data and information interoperable along the different organism of the provincial government named it “technological perspective”. In this way, the data interchange was implemented from the idea of software as service [25] allowing that any authorized organism access to the data in direct and automatic way without intermediaries.

Finally, the three perspectives (technological, business and functional) were visually coordinated through navigational maps jointly with the authorities.

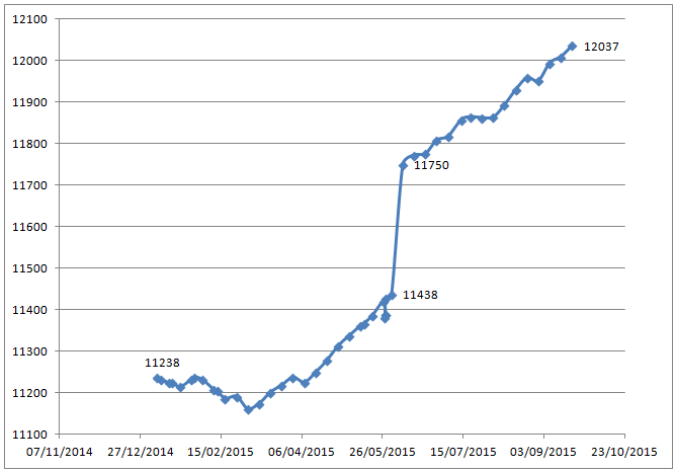

Fig. 3. AIoHLP - Evolution of the volume of Registered Families

In figure 3 we can appreciate the evolution of the volume of registered families between November of 2014 and October of 2015. As you can see, the rate of increasing is really high and it is far of to be stabilized. Nowadays, this “simple” metric (among others) in the command board allows them define the works plan and make the projection of the demand to the future.

As obtained benefits, the AIHLP today is able to make decision guided by data, monitoring their business processes in real time and being interoperable with any authorized organism which requires the information.

V. Related Works

The idea related with the requirement derivation from business process models is widely shared. Herden and others [26] propose the requirement derivation from BPMN (Business Process Model and Notation) 2 process model. The idea is similar, but in our case we are oriented to support the decision-making process on the base of a measurement and evaluation framework such as C-INCAMI. This allows us to improve the comparability, consistency and traceability of the measures along the measurement process.

In [27], Kulkarny and others propose an approaching based on the Enterprise Specification Language (ESL) which allow the simulation and the supporting of the data-driven decision making which minimize the human expertise dependence. The idea of simulating from the data is very interesting, but in contrast, our proposal uses the concept of organizational memory and case-based reasoning for modeling the organizational knowledge and for supporting the recommendation in the decision-making process.

VI. Conclusions

In this invited talk we introduced our perspective on the importance of the measurement and evaluation in the data-driven decision making. In terms of data quality and based on the ISO 25.012, we can have aspects depending exclusively on the data, but also on the system or both. In this sense, the data-driven decision making is dependent of the data quality among other associated factors (e.g. governance, etc). A poor data quality, a poor measurement process, possibly implies a poor decision making process. Thus, the Data Management Maturity Model from the CMMI Institute constitutes interesting an alternative at least tfor its consideration.

We frequently can read papers about the benefits of the data-driven decision making, but in this talk we shown the catastrophes associated with the use the decision-making guided by intuition in a real case in the city of Santa Rosa (La Pampa, Argentina). It is interesting for sizing and making tangible the implications positive and negative related to the presence and absence of the data-driven decision making.

We present an application case related to the Autarchic Institute of Housing of La Pampa, in where the business processes are considered as entity under analysis. In this aspect, the measurement and evaluation strategy using C-INCAMI was defined for supporting the decision-making process.

As future work, we will advance on the implementation of different study cases associated with data-driven decision making, limitations and implications.

Acknowledgment

This research is supported by the project 278/16 of the Economics and Law School and the project 09/F068 of the Engineering School.

References

[1] Van Gigch, J, General System Theory (In Spanish), Trillas, Ed. México: Trillas, 1995.

[2] L. Von Bertalanffy, General System Theory. Foundations, Development, Applications, 2nd ed. México: Fondo de Cultura Económica, 2006.

[3] S Vaezi, "Measurement and Evaluating Frameworks in Electronic Government Quality Management," in 2nd International Conference on Theory and Practice of Electronic Governance, Cairo, Egypt., 2008, pp. 160-165.

[4] P Becker, F Papa, and L Olsina, "Enhancing the Conceptual Framework Capability for a Measurement and Evaluation Strategy," in International Conference on Web Engineering, Aalborg, 2013, pp. 104-106.

[5] D. Thakkar, A Hassan, G. Hamann, and P Flora, "A Framework for Measurement Based Performance Modeling," in 7th International Workshop on Software and Performance, Princeton, NJ, USA, 2008, pp. 55-66.

[6] P. Becker, P. Lew, and L. Olsina, "Strategy to Improve Quality for Software Applications: a Process View," in International Conference on Software and Systems Process, Waikiki, Honolulu, 2011, pp. 129-138.

[7] M Martin and M Diván, "Applications of Case Based Organizational Memory Supported by the PAbMM Architecture," Advances in Science, Technology and Engineering Systems Journal, vol. 2, no. 3, pp. 12-23, April 2017.
[8] TodoNoticias. (2017, April) The Hydric Emergency is declared by inundations (In Spanish). Todo Noticias. [Online]. https://tn.com.ar/sociedad/la-pampa-otra-vez-golpeada-por-la-naturaleza-emergencia-hidrica-en-santa-rosa-por-las-lluvias_783559. Last accessed: October 31 of 2017.
[9] I. Todoran, L. Lecornu, A. Khenchaf, and J. Caillec, "A Methodology to Evaluate Important Dimensions of Information Quality in Systems," ACM. Journal of Data and Information Quality, vol. 6, no. 2-3, pp. 11:1--11:23, June 2015.
[10] P. Becker, H. Molina, and L. Olsina, "Measurement and evaluation as a quality driver," Journal Ingénierie des Systèmes d'Information (JISI), vol. 15, no. 6, pp. 33-62, 2010, Special Issue "Quality of Information Systems".
[11] R. Kaplan and D. Norton, "The Balanced Scorecard – Measures That Drive Performance," Harvard Business Review, vol. 70, no. 1, pp. 71-79, 1992.
[12] R. Kaplan and D. Norton, "Using the Balanced Scorecard as a Strategic Management System," Harvard Business Review, vol. 71, no. 1, pp. 75-85, 1996.
[13] V. Basili, G. Caldiera, and D. Rombach, "The Goal Question Metric Approach," in Encyclopedia of Software Engineering.: Wiley, 1994, vol. I, pp. 528-532.
[14] M Diván, "Strategy for Data Stream Processing based on Measurement Metadata (In Spanish)," UNLP, La Plata, Buenos Aires, Argentina., PhD Thesis 2011.
[15] ISO, ISO-IEC 25012:2011. Software Engineering - Software Product Quality Requirements and Evaluation (SQuaRE) - Data Quality Model.: International Organization for Standardization (ISO), 2011.
[16] F Provost and T Fawcett, "Data Science and its Relationship to Big Data and Data-Driven Decision Making," Big Data, vol. 1, no. 1, pp. 51-59, March 2013.
[17] M Sutherland and M Cook, "Data-Driven Smart Cities: A Closer Look at Organizational, Technical and Data Complexities," in 18th Annual International Conference on Digital Government Research, Staten Island, NY, USA, 2017, pp. 471-476.
[18] CMMI Institute. (2017, October) Data Management Maturity Model. Official web site. [Online]. http://cmmiinstitute.com/data-management-maturity. Last accessed: October 31 of 2017.
[19] I Caballero and M Piattini, "CALDEA: a data quality model based on maturity levels," in 3rd International Conference on Quality Software, 2003, pp. 380-387.
[20] M Diván, "Processing Architecture based on Measurement Metadata," in 5th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO), Noida, India, 2016, pp. 6-15.
[21] M. Martín and L. Olsina, "Added Value of Ontologies for Modeling an Organizational Memory," in Building Organizational Memories: Will You Know What You Knew?, J Girard, Ed. USA: IGI Global, 2009, ch. 10, pp. 127-147.
[22] B Dillon, "The Inundation viewed from a fly of the Geographic Institute of the National University of La Pampa," La Arena, April 2017. Last accessed: October 31 of 2017.
[23] SPEM, "Software Process Engineering Meta-Model Specification," Object Management Group (OMG), Ver.2.0, 2008.
[24] V Ávalos Serrano and M Diván, "An Integrated Strategy Based in Processes, Requirements, Measurement and Evaluation, for the Formalization of Necessities in Data Warehouse Projects," in International Workshop on Data Mining with Industrial Applications, Asunción, Paraguay, 2015.
[25] Nitu, "Configurability in SaaS (Software As a Service) Applications," in 2Nd India Software Engineering Conference, Pune, India, 2009, pp. 19-26.
[26] A Herden, P Farias, and A Albuquerque, "An approach based on BPMN to detail use cases," in New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering, K Elleithy and T Sobh, Eds.: Springer International Publishing, 2013, vol. 312, pp. 537-544.
[27] V Kulkarni, S Barat, T Clark, and B Barn, "Using simulation to address intrinsic complexity in multi-modelling of enterprises for decision making," in Proceedings of the Conference on Summer Computer Simulation, Chicago, Illinois, 2015, pp. 1-11.
[28] J James, D Witten, T Hastie, and R Tibshirani, An Introduction to Statistical Learning with Applications in R, 8th ed. New York, United States: Springer Science+Business Media, 2017.
[29] M Ilyas and J Küng, "A comparative analysis of similarity measurement techniques through SimReq framework," in 7th International Conference on Frontiers of Information Technology, Abbottabad, Pakistan, 2009, pp. 47:1--47:6.
[30] J Whisell and C Clarke, "Effective measures for inter-document similarity," in 22nd ACM international conference on Information & Knowledge Management, San Francisco, California, USA, 2013, pp. 1361-1370.
[31] S Metzger, R Schenkel, and M Sydow, "Aspect-Based Similar Entity Search in Semantic Knowledge Graphs with Diversity-Awareness and Relaxation," in IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies, Washington, DC, USA, 2014, pp. 60-69.

View publication stats