A BIBLIOMETRIC REVIEW OF PROCESS MINING IN MANUFACTURING PROCESSES

UMA REVISÃO BIBLIOMÉTRICA DA MINERAÇÃO DE PROCESSOS NOS PROCESSOS DE MANUFATURA

Ramon Gomes da Silva✉

✉*ramongs1406@gmail.com*

ABSTRACT. The process mining is essential when it is needed to know what is happening in the execution of the processes, even more when it is manufacturing processes. Also, there is a lack of studies on process mining in manufacturing processes, one reason is because the process mining still a young field of study. Therefore, this paper presents a bibliometric review of a history of the process mining techniques applied in manufacturing processes. It was used, as methodology, a systematic review of the literature, by consulting databases from Web of Science and Scopus, using filters to generate a sample that would match with the scope of the research. As results, it was made some analysis, which presents the identification of the years, authors, research source, countries, number of citations, keywords cloud and industry type used in the cases studied. Also, some graphics, tables and information are presented on this paper.

Keywords: Process mining, Manufacturing, Review, Bibliometric analysis.

RESUMO. A mineração de processos é essencial quando é necessário saber o que está ocorrendo na execução dos processos, ainda mais quando são processos de manufatura. Também, há uma lacuna nos estudos da mineração de processos em processos de manufatura, uma razão é por conta de a mineração ainda ser uma área de estudos recente. Portanto, este trabalho apresenta uma revisão bibliométrica da história das técnicas de mineração de processos aplicada em processos de manufatura. Foi utilizado, como metodologia, uma revisão sistemática da literatura, por meio da consulta das bases de dados *Web of Science* e *Scopus*, usando filtros para gerar uma amostra que correspondesse com o escopo da pesquisa. Como resultados, foram realizadas algumas análises, as quais apresentaram a identificação dos anos, autores, fontes de pesquisa, países, número de citações, nuvem de palavras chaves e o tipo de indústria usados nos estudos de caso. Ainda, alguns gráficos, tabelas e informações são apresentados neste trabalho.

Palavras-chave: Mineração de processos, Manufatura, Revisão, Análise bibliométrica.
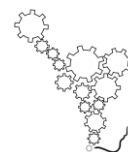
1 INTRODUCTION

The manufacturing industry has a big role on world economy and development. In manufacturing processes, as in a production line, it is essential keep the flow running and know what is happening in each sector of the factory. Hence, process mining techniques are needed to get information about what is really happening in the execution of a process, and not what the people think it is   happening (VAN DER AALST; WEIJTERS; MARUSTER, 2004). Typically, these techniques use the log files that collect information about the events detected and stored by the information system in which the process has been performed (VÁZQUEZ-BARREIROS; MUCIENTES; LAMA, 2015).

Therefore, this paper aims to present a bibliometric review of a history of the process mining techniques applied in manufacturing processes. The contribution of this study is to show who are the principal authors that work on this area, what are the journals or conferences more likely to this research, what are the principal studies about it, etc. In addition, the information found could show if there are lacks on this kind of research, leading future researchers where are the unexplored areas.

This article is structured in five sections. In section 1, an introduction; in the section 2, a theoretical background; in the section 3 the description of the methodology of the samples definition and the analysis techniques; in the section 4, the results; and in the section 5, the final considerations and suggestion for future researches. Also, the references and an appendix at the end.

2 BACKGROUND

Process mining is a recent subject area of research that started to be explored in the last decade. The process mining techniques is used on different areas, such as, manufacturing (YANG et al., 2014), business (NGUYEN et al., 2017), services (VAN DER AALST, 2013), healthcare (REBUGE; FERREIRA, 2012), and others processes.

Furthermore, reinforcing the idea of process mining diversity, the process mining techniques provide new means to improve processes in a variety of application domains (VAN DER AALST et al., 2012). This diversity is due the fact that process mining aims to discover, monitor, and improve real processes by extracting knowledge from event logs readily available in today's information systems (VAN DER AALST, 2013).

To better understanding of what process mining does, the software runs on machines that can log user behavior, which are stored in a so-called event log (VAN GENUCHTEN et al., 2014). Therefore, process mining aims to extract useful process-related information from those event logs.

Moreover, some authors divide the process mining into three types: (1) discovery, (2) conformance, and (3) enhancement (VAN DER AALST, 2013; VAN DER AALST; DUSTDAR, 2012; VAN GENUCHTEN et al., 2014), see Figure 1.

FIGURE 1 – THE THREE BASIC TYPES OF PROCESS MINING
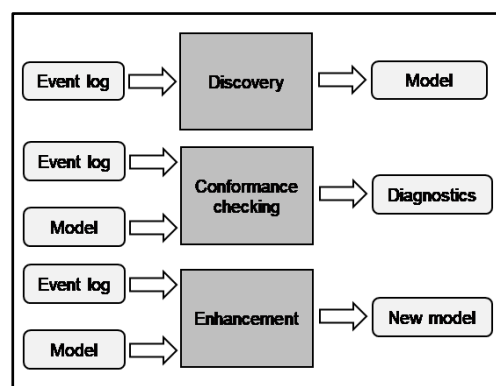
SOURCE: ADAPTED FROM (VAN DER AALST; DUSTDAR, 2012)

3 METHODOLOGY

The bibliometric analysis is defined as an investigation technique that aims to analyze the size, increasing and distribution of the bibliography in a certain knowledge field (ENDLER et al., 2016; PILKINGTON; MEREDITH, 2009).

Citation analysis and content analysis are commonly used bibliometric methods (SUN; WANG; HO, 2012). Bibliometrics have wide applications in various areas to elevate research performance or assess the research trends by investigating the publication characteristics, such as authorship, sources, subjects, geographical origins, and citations (SMALL, 2003).

Therefore, knowing the concepts of bibliometric analysis and the objective of this study on presenting a bibliometric review of a history of the process mining techniques applied in manufacturing processes, this paper followed three steps for the progress of the research, seen on Figure 2.

FIGURE 2 – RESEARCH PROCESS
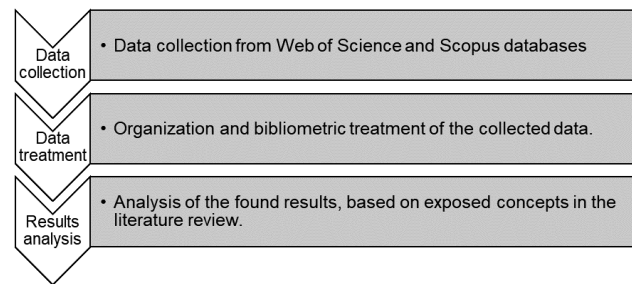
SOURCE: Author (2017)

3.1 DATA COLLECTION

In Table 1 shows the research protocol resumé used on this study. The sample of the articles was collected following the steps detailed below.

TABLE 1 – RESEARCH PROTOCOL

Filters	Entrance data	Databases results		
		Scopus	Web of Science	Total
Topics	"process mining" and "manufact*" or "production"	118	31	149
Source type	Journal and Conference Proceedings	80	31	111
Language	English	73	30	103
Duplicates	29 excluded	-	-	74
Non-relevant	39 excluded	-	-	35

SOURCE: Author (2017)

Two databases were chosen for this work: Scopus and Web of Science, for the reason that both groups a sufficient amount of papers. The first entrée data were the topics "process mining", "manufact*" and "production", resulting in 118 and 31 papers, in Scopus and Web of Science, respectively.

Next, a second filter was to reduce the initial sample according to their source, which had to be from journals and conference proceedings only. That resulted in 80 and 31 articles, following the same order.

Furthermore, the language restriction was applied to determine the articles that could be easily read worldwide. For this, the English was chosen as language restriction and reduced the quantities to 73 papers in Scopus and 30 papers in Web of Science.

Consequently, there were some articles duplicating when both databases results were gathered. 29 articles were excluded as duplicates, remaining the total of 74 papers.

For the last filtering, exclusion of non-relevant articles, some criteria were considered, they are:

- Papers that manufacturing was not the research core;
- Papers that "process mining" was not considered;
- Papers that was about ore mining.

Hence, after the last exclusion, the total amount of papers was 35 articles with no restriction of publication year.

3.2 DATA TREATMENT AND ORGANIZATION

The 35 articles in the sample were cataloged and analyzed the quantities of papers by year; by authors; by research source; by country of origin; by number of citations for each database; by type of industry studied in the paper; and a keywords cloud was plotted.

In addition, to organize the article, *Mendeley Software* (a reference manager) was used to support the organization and filtering the articles according with the necessity, once it showed all the information needed. Moreover, Excel spreadsheets were used in order to generate the graphics, tables, figures, etc. For world map representation, a *Google Maps* tool was utilized to stamp the countries.

4 RESULT ANALYSIS

The first analysis aims to identify the tendencies of increasing or decreasing of the approach about the usage of the process mining techniques in manufacturing processes on the past years, presented on Figure 3.

In fact, the process mining applied on manufacturing processes is very recent, once the first article was published in 2006 by Trappey, Lu and Chen. This is due the fact the process mining is a relative young research discipline (VAN DER AALST, 2016).

FIGURE 3 – NUMBER OF ARTICLES PUBLISHED BY YEAR
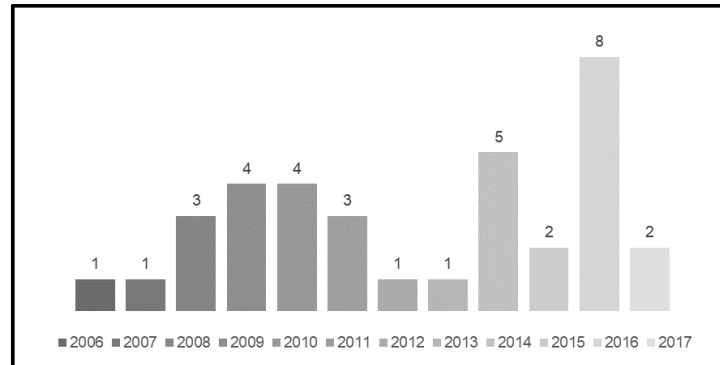
SOURCE: Author (2017)

Moreover, the second analysis is concerning to identify the authors by their publications quantity in the area. In those 35 articles analyzed, there are a total amount of 89 authors, most of them participated in 1 article only.

On the graphic, presented in Figure 4, it is listed all the authors with 2 or more publications, which resulted in 18 authors in total, in knowledge: 4 authors with 4 papers, 3 authors with 3 papers and 11 authors with 2 papers. Also, some authors did publications in partnership, e.g. P. Viale with J. Pinaton, and A. Rozinat with W. van der Aalst.

FIGURE 4 – NUMBER OF PUBLICATIONS BY AUTHOR
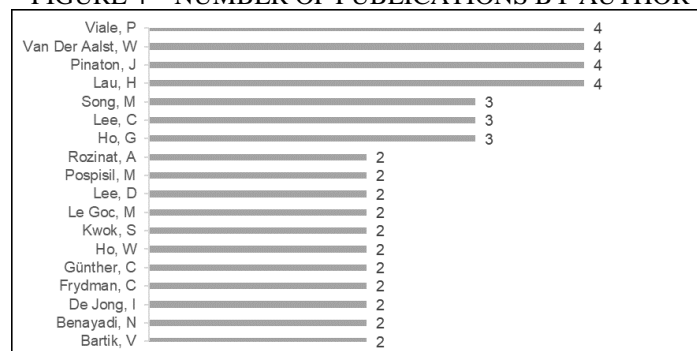
SOURCE: Author (2017)

The third analysis was based on the source type of the articles, to know which journal or conference proceedings have more relevance in the area. The results show that the articles are distributed in various sources, 30 sources in total. In Figure 5 we can observe that only 4 sources have 2 articles published in them, the rest of them have 1 article each.
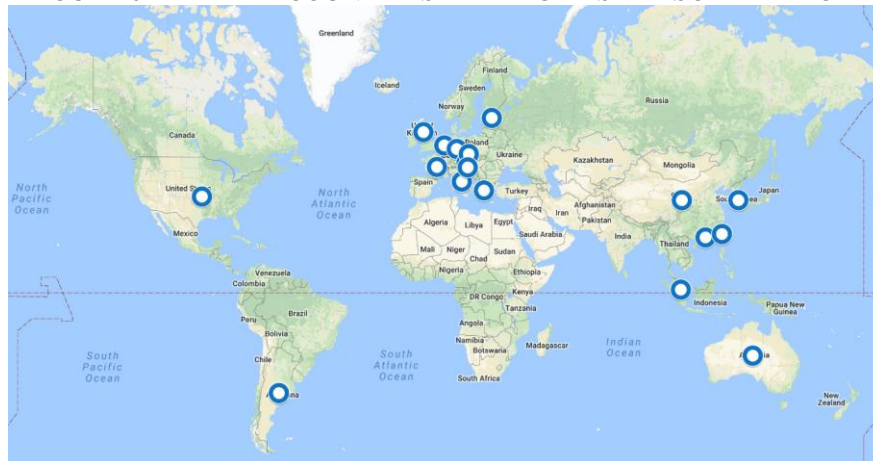
FIGURE 5 – NUMBER OF PUBLICATIONS BY SOURCE
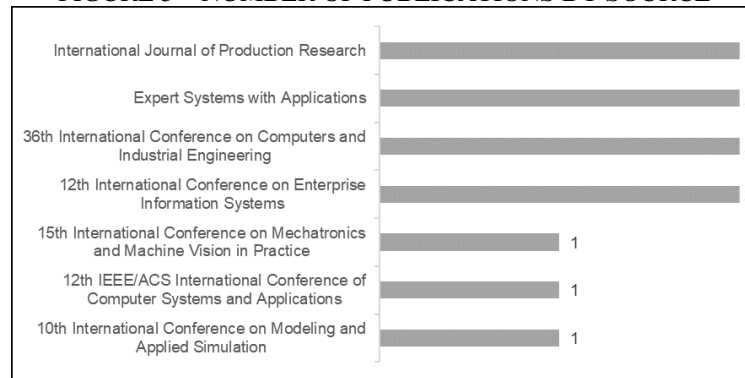
SOURCE: Author (2017)

Next, a classification of the papers was made in order to visualize their distribution by country. That resulted the Figure 6 and 7 which shows on the world map all the countries that did some contribution, and the graph below is the ranking of countries with more publications.
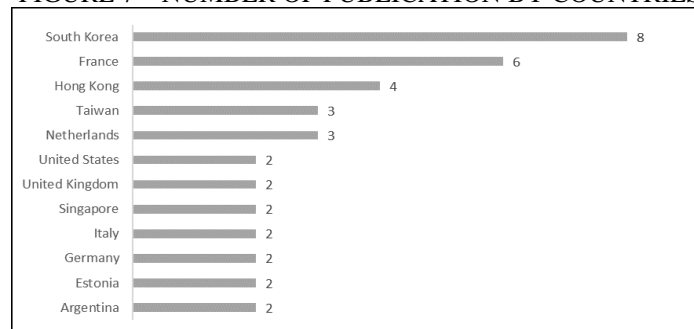
FIGURE 6 – ALL THE COUNTRIES THAT PUBLISHED SOME ARTICLE

SOURCE: Author (2017)

South Korea and France stood out with 8 and 6 articles published, respectively. Also, the graph shows only the countries with at least 2 papers.

FIGURE 7 – NUMBER OF PUBLICATION BY COUNTRIES

SOURCE: Author (2017)

Furthermore, in relation of the relevance of the papers analyzed, the Tables 2 and 3 show the most cited articles in each database, Scopus and Web of Science. For both databases (ROZINAT et al., 2009) and (LAU et al., 2009) are the most cited articles.

TABLE 2 – TOP 10 MOST CITED ARTICLES ON SCOPUS

Authors	Year	Title	# of Citations
Rozinat A., de Jong I.S.M., Günther C.W., van der Aalst W.M.P.	2009	Process mining applied to the test process of wafer scanners in ASML	55
Lau H.C.W., Ho G.T.S., Chu K.F., Ho W., Lee C.K.M.	2009	Development of an intelligent quality management system using fuzzy association rules	28
Rozinat A., De Jong I.S.M., Günther C.W., Van Der Aalst W.M.P.	2009	Conformance analysis of ASML's test process	6
Lee S.-K., Kim B., Huh M., Cho S., Park S., Lee D.	2013	Mining transportation logs for understanding the after-assembly block manufacturing process in the shipbuilding industry	5
Viale P., Frydman C., Pinaton J.	2011	New methodology for modeling large scale manufacturing process: Using process mining methods and experts' knowledge	5
Park J., Lee D., Zhu J.	2014	An integrated approach for ship block manufacturing process performance evaluation: Case from a Korean shipbuilding company	4
Yang H., Park M., Cho M., Song M., Kim S.	2014	A system architecture for manufacturing process analysis based on big data and process mining techniques	4
Jo H., Noh S.D., Cho Y.	2014	An agile operations management system for green factory	3
Cuzzocrea A., Folino F., Pontieri L.	2010	Effective analysis of flexible collaboration processes by way of abstraction and mining techniques	3
Viale P., Benayadi N., Le Goc M., Pinaton J.	2010	Modeling large scale manufacturing process from timed data: Using the TOM4L approach and sequence alignment information for modeling STMicroelectronics' production processes	2

SOURCE: Author (2017)

TABLE 3 – TOP 10 MOST CITED ARTICLES ON WEB OF SCIENCE

Authors	Year	Title	# of Citations
Rozinat, A.; de Jong, I. S. M.; Gunther, C. W.; et al.	2009	Process Mining Applied to the Test Process of Wafer Scanners in ASML	31
Lau, H. C. W.; Ho, G. T. S.; Chu, K. F.; et al.	2009	Development of an intelligent quality management system using fuzzy association rules	21
van der Aalst, Wil M. P.	2010	Process Discovery: Capturing the Invisible	19
Park, Jaehun; Lee, Dongha; Zhu, Joe	2014	An integrated approach for ship block manufacturing process performance evaluation: Case from a Korean shipbuilding company	5
Jo, Hyunjei; Noh, Sang Do; Cho, Yongju	2014	An Agile Operations Management System for Green Factory	3
Yang, Hanna; Park, Minjeong; Cho, Minsu; et al.	2014	A System Architecture for Manufacturing Process Analysis based on Big Data and Process Mining Techniques	3
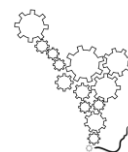
| Lee, Seung-kyung; Kim, Bongseok; Huh, Minhoe; et al. | 2013 | Mining transportation logs for understanding the after-assembly block manufacturing process in the shipbuilding industry | 3 |
Leyer, Michael; Moormann, Juergen	2015	Comparing concepts for shop floor control of information-processing services in a job shop setting: a case from the financial services sector	1
Viale, Pamela; Frydman, Claudia; Pinaton, Jacques	2011	A Process Mining Approach Based on Trace Alignment Information for Very Large Sequential Processes with Duplicated Activities	1
Ho, G. T. S.; Lau, H. C. W.; Kwok, S. K.; et al.	2009	Development of a co-operative distributed process mining system for quality assurance	1

SOURCE: Author (2017)

When crossing information from Tables 2 and 3 with the Figure 4 – the most cited articles with the authors with more publications – it is possible to realize the relevance of some authors (e.g. van der Aalst W., Rozinat A., Lau H.) and their importance in this research area, i.e. which authors are reference in the subject.

In addition, using a cloud of keywords, it was possible to identify the principal keywords used on the all 35 articles analyzed. In Figure 8, the cloud shows the 20 most frequent keywords, where the larger the word, the more it is frequent.

Therefore, the words "process", "mining" and "manufacturing" are the more frequent, which means that the articles of the sample matches with the research made to achieve that sample, then validating the sample. Moreover, other words call the attention – such as "quality", "model", "systems" – this shows that the articles are exploring other subjects beyond the manufacturing.

FIGURE 8 – CLOUD OF THE MOST CITED KEYWORDS BY THE AUTHORS
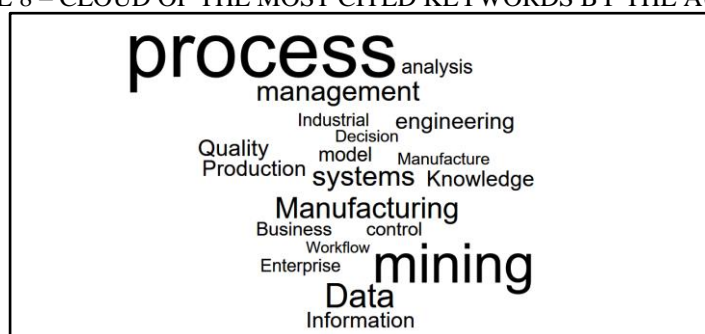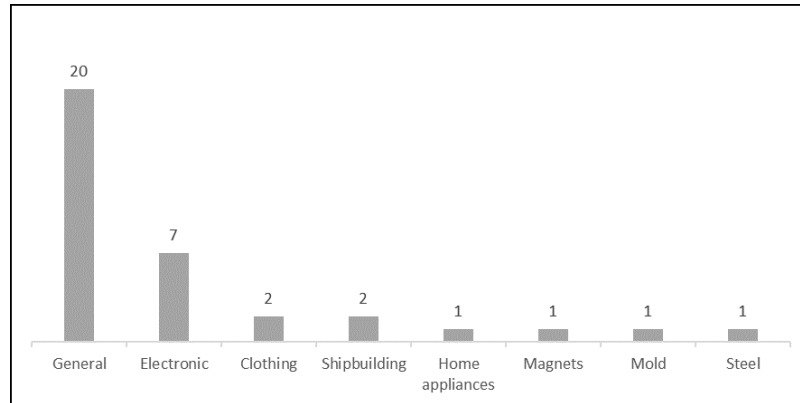

SOURCE: Author (2017)

Last, an analysis was made based on the type of industry that the articles applied their research. As seen in Figure 9, most part of the papers did their research using generic

manufacturing systems, i.e. there are none specification the type of the industry, therefore the research fit in any type of manufacturing industry, theoretically.

FIGURE 9 – CLASSIFICATION OF THE ARTICLES BY THE TYPE OF INDUSTRY

SOURCE: Author (2017)

Those generic articles were classified as general industry, representing the manufacturing as whole. Also, other industry that highlighted was the electronic industry, which includes manufacturing of microchips, computers, electronic equipment, etc.

5 CONCLUSIONS

This study showed that the research in process mining applied in manufacturing processes has a variation over the years, some years its increased, other decreased. The highest point was reached in 2016 with 8 articles, until the moment that this study was made. Also, A. Trappey, T. Lu and R. Chen were the first research to work on this in 2006.

Also, the number of authors that have some contribution to the area has a total of 89 different researches, however 71 of them contributed with 1 publication only, and most of them did some partnership between themselves. About the distribution in relation with the source, it showed to be more dispersed, once that over the 30 sources (journals and conference proceedings), only 4 of them had 2 papers published on them.

In total of 19 countries, South Korea leads the ranking with 8 papers, followed by France with 6 papers, and Hong Kong with 4 papers, as top 3 countries with more publications. In addition, only 7 countries had 1 publication each.

Furthermore, crossing the information about the number of publications by author and the number of citations of the articles, it is possible to affirm that A. Rozinat and W. van der
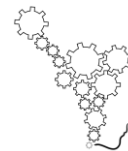
Aalst are the principal researchers and references on the area studied. In both databases, the paper of Rozinat et al. in 2009 was the most cited paper, 55 times in Scopus and 31 in Web of Science.

Moreover, the last analysis about the industries that were used on the studies on the papers analyzed, around 57% of the articles used general examples of manufacturing processes to present their ideas. 15 articles used defined types of industry, such as electronic, clothing, steel, and others.

Finally, because of process mining be a recent field of study, there are no much variability on the articles, this is shown by the size of the sample found, however, this shows an opportunity for future research to explore areas least explored recently. Also, another opportunity is use other databases and check if the results maintain the same pattern or shows some variation. Last, the use of other tools of bibliometric review non-used on this paper, such as co-citation network, is encouraged.

REFERENCES

ENDLER, K. D. et al. Lean seis sigma: uma contribuição bibliométrica dos últimos 15 anos. **Revista Produção Online**, v. 16, n. 2, p. 575, 15 jun. 2016.

LAU, H. C. W. et al. Development of an intelligent quality management system using fuzzy association rules. **Expert Systems with Applications**, v. 36, n. 2 PART 1, p. 1801–1815, 2009.

NGUYEN, H. et al. **Stage-based business process mining**. (F. X. et al., Eds.)Forum and Doctoral Consortium Papers Presented at the 29th International Conference on Advanced Information Systems Engineering, CAiSE-Forum-DC 2017. **Anais**...Queensland University of Technology, Australia: CEUR-WS, 2017

PILKINGTON, A.; MEREDITH, J. The evolution of the intellectual structure of operations management—1980–2006: A citation/co-citation analysis. **Journal of Operations Management**, v. 27, n. 3, p. 185–202, jun. 2009.

REBUGE, Á.; FERREIRA, D. R. Business process analysis in healthcare environments: A methodology based on process mining. **Information Systems**, v. 37, n. 2, p. 99–116, abr. 2012.
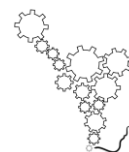
ROZINAT, A. et al. Process mining applied to the test process of wafer scanners in ASML. **IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews**, v. 39, n. 4, p. 474–479, 2009.

SMALL, H. Paradigms, citations, and maps of science: A personal history. **Journal of the American Society for Information Science and Technology**, v. 54, n. 5, p. 394–399, mar. 2003.

SUN, J.; WANG, M.-H.; HO, Y.-S. A historical review and bibliometric analysis of research on estuary pollution. **Marine Pollution Bulletin**, v. 64, n. 1, p. 13–21, jan. 2012.

VAN DER AALST, W. et al. Process Mining Manifesto. In: **Business Process Management Workshops**. [s.l: s.n.]. p. 169–194.

VAN DER AALST, W. Service Mining: Using Process Mining to Discover, Check, and Improve Service Behavior. **IEEE Transactions on Services Computing**, v. 6, n. 4, p. 525–535, out. 2013.

VAN DER AALST, W. M. P. **Process Mining: Data Science in Action**. [s.l.] Springer Berlin Heidelberg, 2016.

VAN DER AALST, W. M. P.; DUSTDAR, S. Process Mining Put into Context. **IEEE Internet Computing**, v. 16, n. 1, p. 82–86, jan. 2012.

VAN DER AALST, W.; WEIJTERS, T.; MARUSTER, L. Workflow mining: discovering process models from event logs. **IEEE Transactions on Knowledge and Data Engineering**, v. 16, n. 9, p. 1128–1142, set. 2004.

VAN GENUCHTEN, M. et al. Is Your Upgrade Worth It? Process Mining Can Tell. **IEEE Software**, v. 31, n. 5, p. 94–100, set. 2014.

VÁZQUEZ-BARREIROS, B.; MUCIENTES, M.; LAMA, M. ProDiGen: Mining complete, precise and minimal structure process models with a genetic algorithm. **Information Sciences**, v. 294, p. 315–333, fev. 2015.

YANG, H. et al. **A system architecture for manufacturing process analysis based on big data and process mining techniques**. (C. W. et al., Eds.)2nd IEEE International Conference on Big Data, IEEE Big Data 2014. **Anais**...School of Business Administration, Ulsan National Institute of Science and Technology, Ulsan, South Korea: Institute of Electrical and Electronics Engineers Inc., 2014