Knowledge Extraction in Organizations and Services Through Process Modeling: A Bibliometric Analysis

Kátia da Silva Vianna Menezes
Master's in Systems Applied to Engineering and Management, Instituto Federal Fluminense (IFF/SAEG), Campos dos Goytacazes, Rio de Janeiro, Brazil.
Administrator and Planning Analyst, Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), Campos dos Goytacazes, Rio de Janeiro, Brazil.
http://lattes.cnpq.br/1678618016099509
https://orcid.org/0000-0001-8991-8245
katia@uenf.br

Gabriela Martins Cipriano
Master's in Systems Applied to Engineering and Management, Instituto Federal Fluminense (IFF/SAEG), Campos dos Goytacazes, Rio de Janeiro, Brazil.
Logistics Operations Analyst, Suzano S.A., Aracruz, Espírito Santo, Brazil.
http://lattes.cnpq.br/7010906457846900
gmccipriano@gmail.com

Gregório Jean Varvakis Rados
PhD in Manufacturing Engineering, Loughborough University of Technology, England, United Kingdom.
Professor, Universidade Federal de Santa Catarina, Graduate Program in Knowledge Engineering and Management (UFSC/EGC), Florianópolis, Santa Catarina, Brazil.
http://lattes.cnpq.br/8153758173673961
https://orcid.org/0000-0003-2576-4835
g.varvakis@ufsc.br

Alexandre Augusto Biz
PhD in Knowledge Engineering and Management, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, Brazil.
Professor, Universidade Federal de Santa Catarina, Graduate Program in Knowledge Engineering and Management (UFSC/EGC), Florianópolis, Santa Catarina, Brazil.
http://lattes.cnpq.br/045636273321595
https://orcid.org/0000-0003-3235-9328
alexandre.biz@ufsc.br

Simone Vasconcelos Silva
PhD in Computer Science, Universidade Federal Fluminense (UFF), Niterói, Rio de Janeiro, Brazil.
Full Professor, Instituto Federal Fluminense, Graduate Program in Systems Applied to Engineering and Management (IFF/SAEG), Campos dos Goytacazes, Rio de Janeiro, Brazil.
http://lattes.cnpq.br/5602424371704966
https://orcid.org/0000-0002-5994-6840
simonevs@iff.edu.br

Submitted: 12/29/2022. Approved: 06/10/2025. Published: 20/10/2025

ABSTRACT

An organization's capacity for continuous innovation is closely linked to how effectively it mobilizes existing knowledge to promote organizational learning. This article addresses the following research question through a bibliometric study: Which technique is most frequently used for knowledge extraction based on business process modeling? The bibliometric analysis identified meaningful connections among the most cited authors, seminal publications, leading sources, and countries contributing most significantly to the field. It also examines the main methodological approaches employed. The sample consisted of 91 studies drawn from 65 sources indexed in the Scopus and Web of Science databases. Since 2004, scientific production on the topic has increased steadily, particularly in China, the United States, and Germany. The years 2011 and 2014 were the most productive, suggesting that the topic remains relatively recent within academic literature. These findings point to opportunities for future research that may involve broader bibliometric analyses, including additional metrics and databases, or systematic literature reviews on knowledge extraction in organizations and services grounded in business process modeling.

Keywords: knowledge management; BPMN; business process management; bibliometrics.

INTRODUCTION

The current organizational landscape, emblematic of Society 5.0, positions knowledge as the new factor of production—an intangible, valuable, and irreplaceable asset. An organization's ability to implement continuous innovation is directly tied to its capacity to mobilize existing knowledge and foster organizational learning (Santos; Rados, 2020).

Awareness regarding the importance of comprehensive management of organizational processes has grown rapidly, primarily fueled by intense global market competition, where only the best-performing companies will survive in the long term (Lizano-Mora *et al*., 2021).

In this context, knowledge sourced externally from clients is seen as a major competitive advantage for companies and institutions. It holds the potential to generate value, support intelligent decision-making regarding products and services, and identify new opportunities (Muniz; Dandolini; Biz, 2021).

Well-executed bibliometric studies can lay the groundwork for advancing knowledge across various fields in meaningful ways, enabling scholars to develop a unique overview, identify research gaps, generate new investigative ideas, and direct their contributions to a specific domain (Donthu *et al*., 2021).

Thus, bibliometric reviews represent one of the most effective approaches to describing the body of literature within a specific scientific field. They indicate trends in productivity, relationships, quality, citations, and other key data through a structured process that systematically describes the documents under investigation (Khatib *et al*., 2021).

This article aims to address the following research question through a bibliometric study: What is the most commonly used technique for knowledge extraction through business process modeling?

The article is organized into six sections: (i) introduction; (ii) literature review on Knowledge Management (KM) in organizations and services and on Business Process Management (BPM); (iii) methodological procedures; (iv) presentation and analysis of results; (v) final considerations; and (vi) references.

LITERATURE REVIEW

Knowledge Management in Organizations and Services

Transforming data into information and knowledge plays a crucial role in fostering innovation and sustaining competitiveness, particularly amid the global technological advancement of organizations. In this regard, intangible resources—such as knowledge and experience, stimulate innovation, creativity, and improved decision-making among stakeholders (Mirafzal *et al*., 2023).

Knowledge Management (KM) is a multidisciplinary field that integrates psychology, epistemology, and cognitive science. Its primary objective is to enable individuals and

organizations to share, develop, apply, and renew knowledge in ways that enhance productivity, stimulate innovation, and strengthen both individual and organizational learning (Bencsik, 2021).

Takeuchi and Nonaka (2008) argue that the creation of new knowledge depends on drawing upon the tacit and highly subjective insights of employees—mobilizing and embedding these insights into technologies. Thus, KM can be understood as the management of structured information directed toward specific objectives to generate tangible benefits.

The service sector benefits substantially from the effective implementation of KM, given its dependence on specialized expertise. In this area, KM can improve efficiency by reducing costs, increasing sales and profits, and reinforcing the sector's capacity to maintain growth and competitiveness (Alharbi, 2024).

Within this perspective, Organizational Knowledge Management (OKM), when examined through a "hard" research approach, operates as a powerful amplifier of information management across knowledge bases. It redefines the conventional use of databases by transforming knowledge into a key competitive resource through the application of information technologies (Santos; Rados, 2020).

Business Process Management (BPM)

BPM is a discipline that integrates management and Information Technology (IT), with the latter supporting a process-based management approach. This approach involves identifying, designing, executing, documenting, measuring, monitoring, and controlling both automated and non-automated business processes, aiming to achieve results aligned with the organization's strategic goals (ABPMP, 2020).

According to Jeston (2018), BPM refers to achieving organizational objectives by improving, managing, and controlling core business processes. This aligns with the definition offered by Harmon (2019), who views BPM as a discipline focused on enhancing organizational efficiency through process management.

Kampik *et al*. (2024) characterize BPM as both a professional discipline and a research field that ensures organizations function as intended and achieve their competitive and social objectives. The authors emphasize that BPM is highly knowledge-intensive, which creates challenges for traditional statistical inference methods in leveraging organizational knowledge, particularly because such knowledge is often unstructured and not easily processed.

Process modeling creates a representation of a process to describe it, whether in a simplified or more detailed manner, depending on the goals of the modeling project. Process models are essential tools that support management, analysis, and the proposal of improvements (ABPMP, 2020).

Through process modeling and the application of Intelligent Business Process Management Suites (iBPMS), organizations can leverage functionalities such as Big Data Analytics, Artificial Intelligence (AI) and Machine Learning, Mobile BPM, Social BPM, Cloud-

Based BPM, Process Mining, Robotic Process Automation (RPA), Predictive Analytics, Internet of Things (IoT), Blockchain, and Dynamic Case Management (ABPMP, 2020). These technologies not only automate and optimize processes but also enable the extraction of organizational knowledge, thereby supporting strategic decision-making (ABPMP, 2020).

Beheshti *et al*. (2023) examined the intersection of Artificial Intelligence (AI), BPM, and Knowledge Management (KM), advocating for a rethinking of BPM as organizations seek to better understand the behavior of their information systems, processes, and services. The emergence of generative AI has further expanded these possibilities, allowing organizations to enhance their processes in ways previously unimaginable.

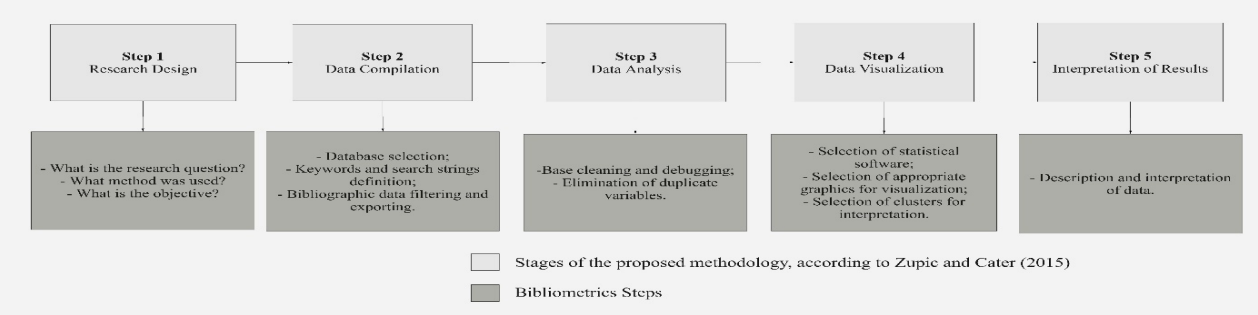
METHODOLOGICAL PROCEDURES

This study adopts a quantitative survey combined with a bibliometric analysis to map the body of knowledge produced on the topic, presenting results through metric indicators and offering analytical insights into the selected theme (Miguel; Gerlin; Da Costa, 2024).

Bibliometric analysis enables the discovery of previously unknown information, the development of indicators for a specific topic, and the rapid, objective identification of relevant relationships (Moura et al., 2017).

The study follows the stages proposed by Zupic and Čater (2015) (**Figure 1**).

Figure 1 – Bibliometric Stages



Source: Author's elaboration, 2022.

Stage 1 was described in the Introduction. Stage 2 involved searching for studies in the *Scopus* and *Web of Science* (WoS) databases, chosen for their peer-reviewed content and their status as two of the most comprehensive bibliographic collections in the scientific community. The search strategy was developed using the keywords and thesauri shown in **Table 1**. The following search string was applied: (*TITLE-ABS-KEY* [*knowledge*] *AND TITLE-ABS-KEY* [*extraction*] *AND TITLE-ABS-KEY* [*servic* OR organization* OR institution**] *AND TITLE-ABS-KEY* [*"business process management" OR BPM OR "process model*"*]).

Bibliographic data were exported in .bib and .txt formats. Although *Google Scholar* indexes a vast number of publications, it was not used because it lacks an Application Programming Interface (API) that allows exporting a set of documents with cited references

(Zupic; Čater, 2015).

Table 1 – Keywords and Thesauri

Topic:	Knowledge Extraction in Organizations and Services Through Process Modeling: A Bibliometric Analysis				
Term:	extraction	knowledge	organization*	service*	business process management
Thesauri:	-	-	institution*	-	BPM
	-	-	-	-	process model*

Source: Author's elaboration, 2022.

Stage 3 emplyed the free and open-source software RStudio (version 2021.09.2 Build 382) to clean and refine the datasets by identifying and removing duplicate records, resulting in a unified database combining *Scopus* and *WoS* outputs.

Subsequently, in Stage 4, the *Biblioshiny* package for *Bibliometrix* was employed. This tool enables quantitative analyses in scientometrics and bibliometrics (Aria; Cuccurullo, 2017) through an intuitive interface that facilitates visualization and configuration of the bibliometric dashboard. *Bibliometrix* was chosen because R offers an open-source environment with robust statistical algorithms and integrated data visualization capabilities.

The bibliometric parameters defined for analysis in Biblioshiny are summarized in **Table 2**.

Table 2 – Bibliometric Data

General	Annual scientific production
Sources	Most relevant sources; Most cited sources; Bradford's Law; Local Source Impact
Authors	Most relevant authors by output; Three-Fields Plot (linking countries, authors, and keywords); Impact Factors (H-index, G-index, M-index); Most relevant affiliations; Scientific production by country
Documents	Most cited works; Referenced Publication Year Spectroscopy (RPYS); TreeMap; Evolution of keyword usage; Word cloud
Conceptual, Intellectual, and Social Structure	Keyword co-occurrence; Network analysis; Co-citation network; Factorial analysis (topic dendrogram); Social structure; Global collaboration map

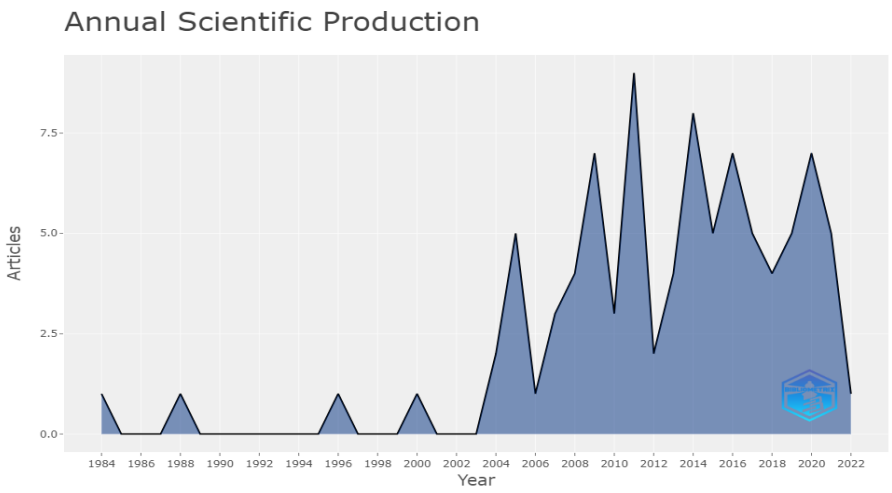
Source: Author's elaboration, 2022.

Data Presentation and Analysis

Applying the proposed search strategy yielded 70 studies in the *Scopus* database and 41 in the *Web of Science* (WoS). After data processing with RStudio software, 20 duplicate publications were removed, resulting in a final sample of 91 studies for analysis.

There was no temporal restrictions applied in the search. The final dataset spans from 1984 to 2022 and includes contributions from 212 authors published across 65 different sources (books, journals, conferences, and others), with only five single-authored works. One publication was identified in 1984 and another in 1988, followed by gaps until 1996 and again until 2000. From 2004 onward, there was a notable increase in annual scholarly output on the subject—likely influenced by the so-called "Third Wave of BPM," which significantly advanced the discipline (Jeston, 2018). The most productive years were 2011 and 2014, with nine and eight articles published, respectively (**Figure 2**)

Figure 2 – Annual Scientific Production



Source: Author's elaboration, 2022.

Source Identification

A total of 65 relevant sources were identified, with the ten most prolific presented in **Table 3**. The Lecture Notes in Computer Science (LNCS) series—including its subseries Lecture Notes in Artificial Intelligence (LNAI) and Lecture Notes in Bioinformatics (LNBI)—accounted for the highest number of publications on the topic (eight articles), followed by Lecture Notes in Business Information Processing, with seven.

Table 3 - Most relevant sources

Sources	Number of articles
Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	8
Lecture Notes in Business Information Processing	7
11th International Multidisciplinary Scientific Geoconference and Expo - Modern Management of Mine Producing Geology and Environmental Protection SGEM 2011	3
Business Process Management Journal	3
International Conference on Information Systems 2011 ICIS 2011	3
Advanced Materials Research	2
Association for Information Systems - 11th Americas Conference on Information Systems AMCIS 2005: A Conference on a Human Scale	2
Data & Knowledge Engineering	2
IEEE Acess	2
IEEE Journal of Biomedical and Health Informatics	2

Source: Author's elaboration, 2022.

Table 4 presents the most cited sources, with LNCS leading by a wide margin (90 citations). LNCS disseminates state-of-the-art research in computer science and is indexed by major databases, including Web of Science, Scopus, Google Scholar, and EI Engineering Index.

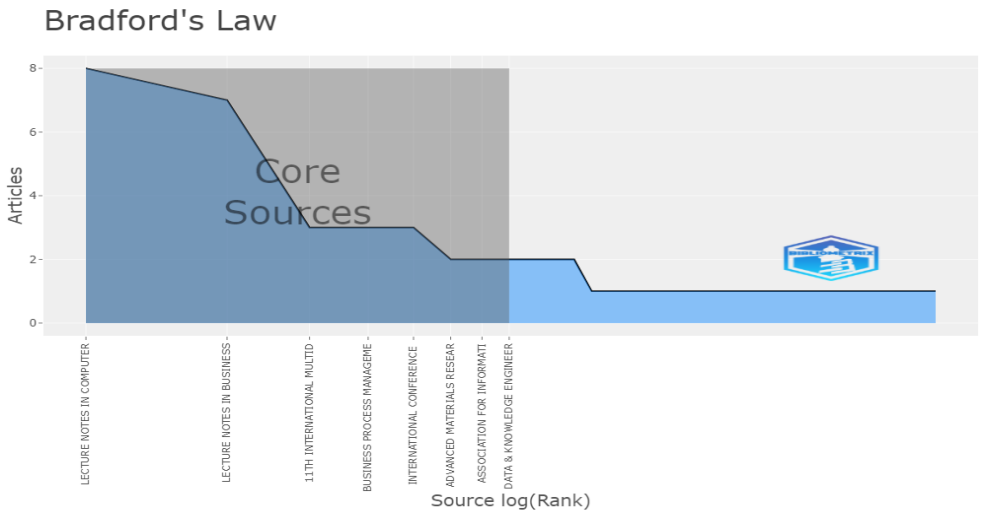
Table 4 - Most Cited Sources

Sources	Citations
Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	90
Information Systems Frontiers	24
Computers in Industry	22
Expert Systems With Applications	20
Lecture Notes in Business Information Processing	19
Decision Support Systems	18
Journal of Machine Learning Research	17
Information Systems Frontiers	13
International Journal of Operations and Production Management	13
Business Process Management Journal	12

Source: Author's elaboration, 2022.

FIGURE 3, based on Bradford's Law, highlights the core sources for the selected terms: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Lecture Notes in Business Information Processing, 11th International Multidisciplinary Scientific Geoconference and Expo – Modern Management of Mine Producing, Geology and Environmental Protection, SGEM 2011, Business Process Management Journal, International Conference on Information Systems 2011, ICIS 2011, Advanced Materials Research e Association for Information Systems – 11th Americas Conference on Information Systems, AMCIS 2005: A Conference on a Human Scale, Data, and Knowledge Engineering. Those sources is where the majority of relevant research is concentrated (Desai; Veras; Gosain, 2018).

FIGURE 3 – Bradford's Law



Source: Author's elaboration, 2022.

FIGURE 4 illustrates the local citation impact of these sources, with Data & Knowledge Engineering, Expert Systems with Applications, and Lecture Notes in Business Information Processing standing out as the most influential.

FIGURE 4 – Local Source Impact

Source: Author's elaboration, 2022.

Author Identification

Amit V. Deokar, from the Manning School of Business (United States), is the most prolific author on the topics of BPM, knowledge extraction, and services, with three publications. **Table 5** lists the ten most productive authors..

Table 5 - Most Relevant Authors

Authors	Articles
DEOKAR, A.	3
AWITI, J.	2
BAIAO, F.	2
BATCHKOVA, I.	2
CALVANESE, D.	2
GONCALVES, J.	2
IVANOVA, T.	2
LI, L.	2
MENDLING, J.	2
MONTALI, M.	2

Source: Author's elaboration, 2022.

The most commonly used metrics for assessing the significance of an author's contributions to the literature are the H-index (Hirsch, 2005), the G-index, and the M-index. The impact scores of the ten authors listed in **Table 4** indicate that Professors Jan Mendling (Humboldt-Universität zu Berlin) and Wil Van Der Aalst (RWTH Aachen University) stand out, particularly regarding the M-index and total citations. Mendling, who began publishing on the subject in 2014, has accumulated 54 citations, an H-index of 2, a G-index of 2, and an M-index of 222. These metrics highlight his relevance in the field, especially considering the relatively short period since his first publication on the topic compared to other highly cited authors who have been contributing since as early as 2005.

Table 6 - Author Impact Factors

Author	H-index	G-index	M-index	Citations	Publications	First Publication Year
MENDLING, J.	2	2	222	54	2	2014
VAN, D. A. W.	2	2	111	51	2	2005
DUSTDAR, S.	1	1	56	37	1	2005
HOFFMANN, T.	1	1	56	37	1	2005
WANG, Y.	1	2	111	36	2	2014
CARON, F.	1	1	111	35	1	2014

GUO, Y.	1	1	111	35	1	2014
HUANG, L.	1	1	111	35	1	2014
VANTHIENEN, J.	1	1	111	35	1	2014
CABANILLAS, C.	1	1	125	33	1	2015

Source: Author's elaboration, 2022.

In terms of institutional affiliation, the Aragon Health Sciences Institute (Spain) leads with four publications, followed by fourteen institutions with two each (**Table 5**).

Table 7 - Most Relevant Affiliations

Afiliations	**Articles**
Aragon Health Sciences Institute (IACS)	4
Brandenburg University of Applied Sciences	2
Dakota State University	2
Eindhoven University of Technology	2
Federal University State Rio De Janeiro - UNIRIO	2
Free University of Bozen-Bolzano	2
Institute Tecnology Buenos Aires	2
Narsee Monjee Institute of Management Studies	2
Neoma Business School	2
Southeast University	2
Technical University of Berlin	2
Université Libre De Bruxelles	2
University of Michigan	2
University of Toronto	2
University of Amsterdam	2
Australian National University	1
Beihang University	1
Beijing Jiaotong University	1
Bournemouth University	1
California State University	1

Source: Author's elaboration, 2022.

Regarding country contributions, China and the United States lead with fifteen publications each, followed by Germany with fourteen. Brazil ranks twelfth among the twenty-eight contributing countries (**Table 8**).

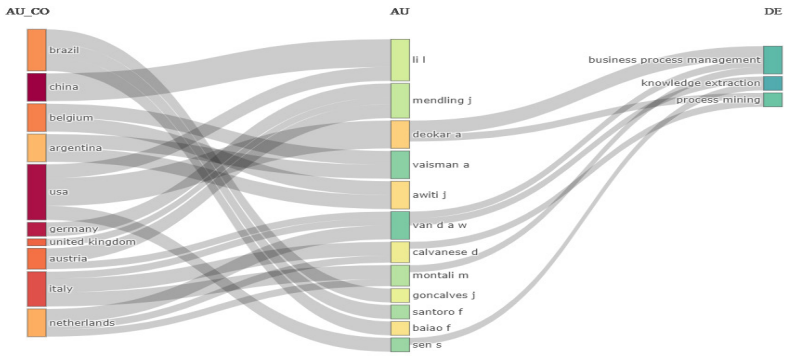
Table 8 - Scientific Output by Countryzzzz

Country	**Nº of Publications**
China	15
Unted Stats	15
Germany	14
France	8
Australia	7
Spain	7
Italy	6
Canada	5
United Kingdom	5
Austria	4
Belgium	4
Brazil	3
India	3
Netherlands	3
Argentina	2

Source: Author's elaboration, 2022.

Figure 5 depicts the connections between countries (Column 1), authors (Column 2), and keywords (Column 3) using a Sankey diagram, also known as a Three-Fields Plot (Riehmann; Hanfler; Froehlich, 2005). Although Brazil appears in the diagram, it shows no connection with the keywords business process management, knowledge extraction, or process mining within the analyzed sample. The countries showing the strongest interactions across the three variables are the United States, Austria, Italy, and the Netherlands.

Figure 5 – Connections Between Authors, Countries, and Keywords (Three-Fields Plot)



Source: Author's elaboration, 2022.

Document Analysis

The most cited article in the dataset is *"Mining of ad-hoc business processes with TeamLog"* by Dustdar et al. (2005), published in *Data and Knowledge Engineering*, with 37 citations. The second most cited is *"Acquiring logistics process intelligence: Methodology and an application for a Chinese bulk port"* by Wang *et al*. (2014), with 35 citations. In third place is *"Mining the Organisational Perspective in Agile Business Processes"* by Schönig *et al*. (2015), with 33 citations. This article explores the extraction of complex resource allocation rules that integrate control flow and organizational perspectives (**Table 9**).

Table 9 - Most Cited Works

Position	Author	Year	DOI	Total Citations	Citations per Year
1	DUSTDAR, S. et al.	2005	10.1016/j.datak.2005.02.002	37	2.0556
2	WANG, Y. et al.	2014	10.1016/j.eswa.2013.07.021	35	3.8889
3	SCHÖNIG, S. et al.	2015	10.1007/978-3-319-19237-6_3	33	4.1250
4	WEIDLICH, M. et al.	2014	10.1109/TKDE.2014.2302306	21	2.3333
5	GUO, L.A. et al.	2017	10.1108/BPMJ-05-2015-0065	16	2.6667

Position	Author	Year	DOI	Total Citations	Citations per Year
6	ROOS, M. et al.	2009	10.1186/1471-2105-10-S10-S9	16	1.1429
7	GEIERHOS, M. et al.	2015	10.5220/0005346002770283	15	1.8750
8	CALVANESE, D. et al.	2016	10.1007/978-3-319-42887-1_12	14	2.2000
9	LI, J.X. et al.	2015	10.1007/s10796-015-9564-3	14	1.7500
10	OKOYE, K. et al.	2014	10.1016/j.procs.2014.08.031	13	1.4444

Source: Author's elaboration, 2022.

Table 10 summarizes the characteristics of these top-cited studies. Notably, 60% employ process mining techniques, and 80% report applications in service-oriented organizations.

Table 10 - Characteristics of the Most Cited Articles

	Country	Institution	Source	Title	Justification	Methodology
1	Áustria	Vienna University of Technology	Data and Knowledge Engineering	Mining of ad-hoc business processes with TeamLog	Difficulty modeling control flow in ad-hoc processes	Application of process mining techniques and tools such as EMiT and MinSoN to analyze ad-hoc processes; use of the TeamLog extraction tool in a banking environment.
2	China	Beijing Jiaotong University	Expert Systems with Applications	Acquiring logistics process intelligence: Methodology and an application for a Chinese bulk port.	Understanding logistics process dynamics to mitigate risks and gain strategic advantage	Case study conducted in a major Chinese port, applying log extraction and preprocessing, followed by exploratory, performance, and conformance analyses.
3	Alemanha	University of	Lecture Notes in Business Information Processing	Mining the Organisational Perspective in Agile Business Processes	Agile processes require explicit integration of the organizational perspective due to the central role of human decisions and specialized knowledge	Proposed a process mining approach to uncover resource knowledge using declarative process models; extracted complex resource allocation rules; applied to university travel approval logs.
4	Alemanha	Humboldt-Universität zu Berlin	IEEE Transactions on Knowledge and Data Engineering	Optimizing Event Pattern Matching Using Business Process Models	Need to leverage business process models for optimizing complex event processing	Extraction of behavioral constraints to rewrite event detection patterns and select execution plans; applied to an insurance company.

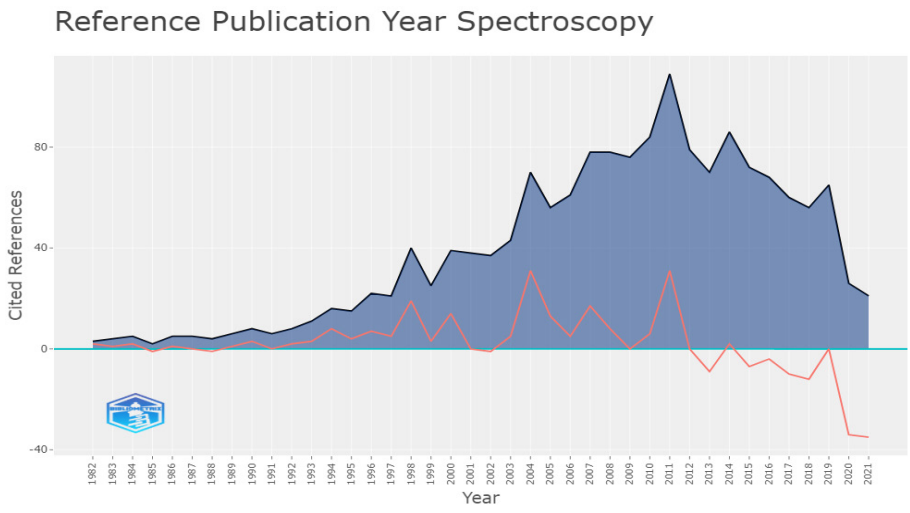
	Country	Institution	Source	Title	Justification	Methodology
5	França	Neoma Business School	Business Process Management Journal	Automated competitor analysis using big data analytics: Evidence from the fitness mobile app business	Many business decisions rely on competitor product analysis	Integration of web crawlers, natural language processing algorithms, and machine learning with data visualization to build a big data–based competitor analysis system; applied to the fitness app industry.
6	Holanda	Informatics Institute, University of Amsterdam	Semantic Web Applications and Tools for Life Sciences	Structuring and extracting knowledge for the support of hypothesis generation in molecular biology	Challenge of identifying relevant information among millions of PubMed publications	Combined use of Semantic Web and information extraction and retrieval techniques to make knowledge available for computational analysis and inference in scientific publications.
7	Alemanha	University of Paderborn, Germany	Proceedings of the International Conference on Agents and Artificial Intelligence	What Did You Mean? - Facing the Challenges of User-generated Software Requirements	Need to assist service consumers with limited expertise in articulating software requirements in natural language	Ontology-based requirement extraction and similarity retrieval from marketplace applications; applied to logistics-related onboarding processes.
8	Itália	Free University of Bozen-Bolzano	Lecture Notes in Business Information Processing	Ontology-Driven Extraction of Event Logs from Relational Databases	In many datasets, event logs are implicitly stored in legacy systems using relational databases	Development of a framework to extract XES event logs from relational databases, enabling compatibility with process mining tools; applied to a conference submission portal.
9	Estados Unidos	Oregon State University	Information Systems Frontiers	An intelligent approach to data extraction and task identification for process mining	The main challenge in business process mining is extracting relevant data from numerous event logs in databases	Document analysis using text mining techniques to identify key tables for process mining based on financial transaction data from a company.
	Reino Unido	University of East London	Procedia Computer Science	A Semantic rule-based approach supported by process mining for personalised adaptive learning	The gap in extracting useful patterns from data sources for knowledge generation	Combination of semantic rule-based modeling and process mining techniques to identify personalized adaptive learning paths.

Source: Author's elaboration, 2022.

These studies encompass diverse organizational contexts – banks, universities, insurance companies, logistics firms, and port management – demonstrating the broad applicability of knowledge extraction through approaches grounded in process management, knowledge management, and information technologies.

The Referenced Publication Year Spectroscopy (RPYS) analysis (**Figure 6**) identifies breakthrough years—1998, 2004, 2011, 2014, and 2019 – marking periods of foundational or disruptive publications within the analyzed timeframe.

Figure 6 – *Referenced Publication Year Spectroscopy*

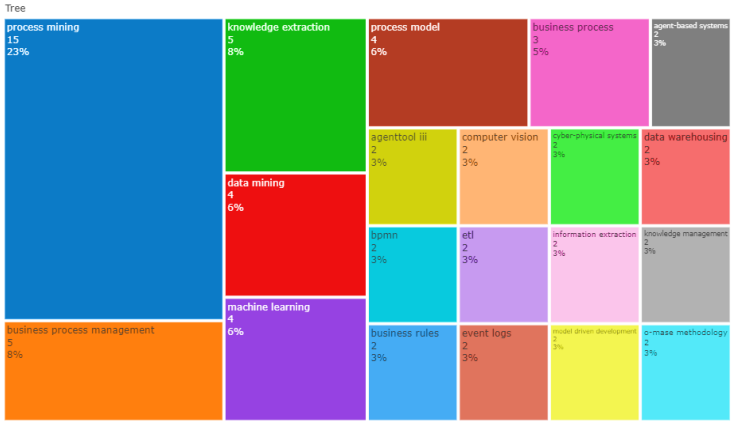


Source: Author's elaboration, 2022.

The TreeMap in **Figure 7** displays the 20 most frequently used keywords, allowing for thematic mapping of the research field. The term process mining appears in 23% of studies, followed by business process management and knowledge extraction (8% each), while data mining, machine learning, and process model each occur in 6%.

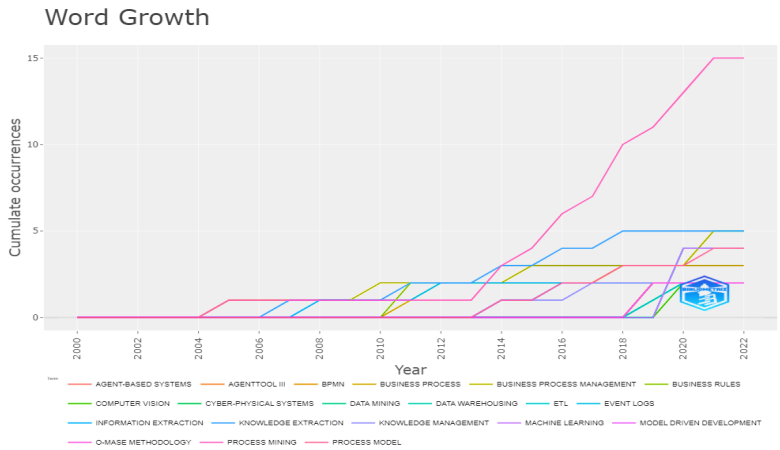
Figure 8 illustrates the evolution of keyword usage shown in **Figure 7**, demonstrating that process mining experienced the most significant growth – far exceeding the average frequency of other terms.

FIGURE 7 – *TreeMap*



Source: Author's elaboration, 2022.

FIGURE 8 – Keyword Usage Evolution



Source: Author's elaboration, 2022.

According to Calvanese *et al*. (2016), process mining is an expanding discipline that seeks to discover, monitor, and improve real processes by extracting knowledge from event logs that capture actual system executions.

Van der Aalst (2022) emphasizes that process mining enhances operational processes through the systematic use of event data – identifying bottlenecks, detecting deviations, and enabling automation or the elimination of repetitive tasks.

FIGURE 9 presents a word cloud of the 50 most frequent author keywords, reinforcing the prominence of *knowledge extraction*, *business process management*, *process model*, *data mining*, and *machine learning*.

FIGURE 9 – Word Cloud of Keywords

Source: Author's elaboration, 2022.

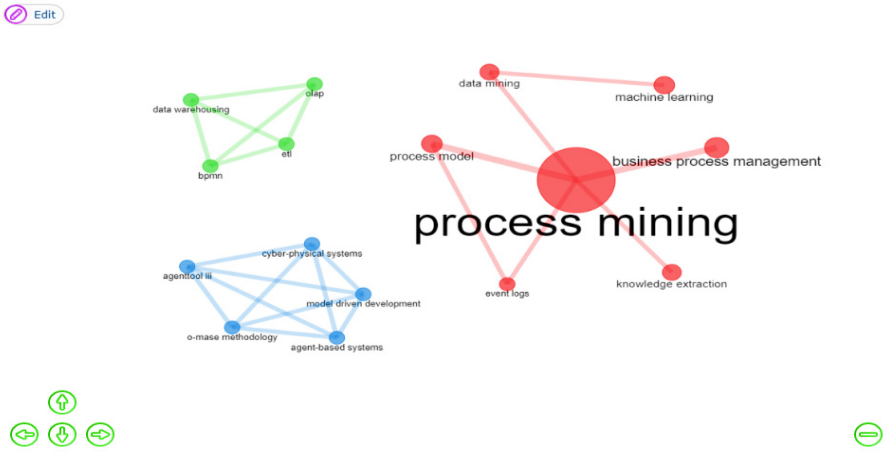
Conceptual, Intellectual, and Social Structures

According to Aria and Cuccurullo (2017), the structure of knowledge can be analyzed across three domains: conceptual, intellectual, and social.

Conceptual Structure

The keyword co-occurrence network is a frequency matrix in which the elements represent co-citations, shared references, co-authorships, or word co-occurrences (Zupic; Čater, 2015). **FIGURE 10** illustrates the co-occurrence of author-defined keywords, highlighting the red cluster, where process mining occupies a central position in the network. It is closely associated with the following themes: data mining, machine learning, process modeling, business process management, knowledge extraction, and event logs.

FIGURE 10 – Keyword Co-occurrence

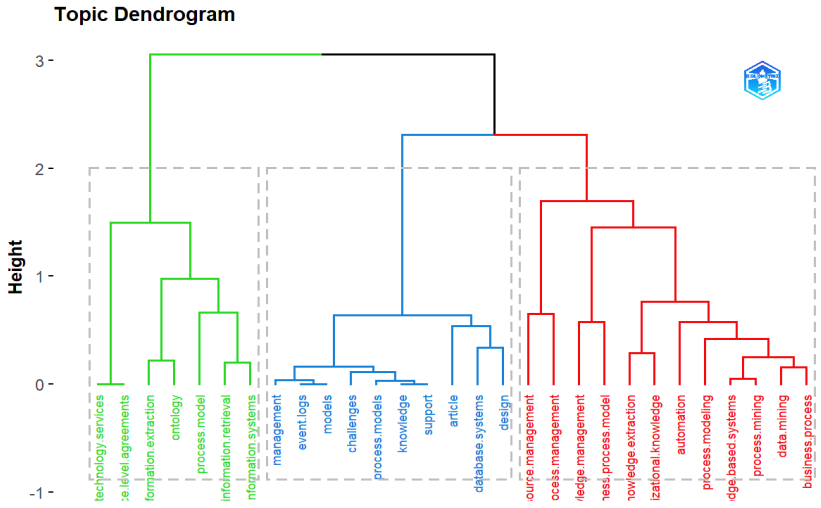


Source: Author's elaboration, 2022.

In bibliometric analysis, Hierarchical Cluster Analysis (HCA) is widely used to identify subgroups, producing a dendrogram of topics based on the similarity among analyzed items (Zupic; Čater, 2015). Applying HCA through the Multiple Correspondence Analysis (MCA) method to the *KeyWords Plus* field yielded 20 terms, grouped into three clusters. The resulting dendrogram (**Figure 11**) demonstrates clear dimensional separation.

KeyWords Plus terms are derived from the titles of articles cited by the author of an indexed work, even if those terms do not appear in the title of the indexed article itself (Clarivate, 2022). **Cluster 3 (Red) i**ncludes terms related to data and process mining, organizational knowledge management, and business processes; it is the most representative cluster. As for c**luster 1 (Green), it c**omprises terms primarily associated with information systems and ontologies.

Figure 11 – Topic Dendrogram

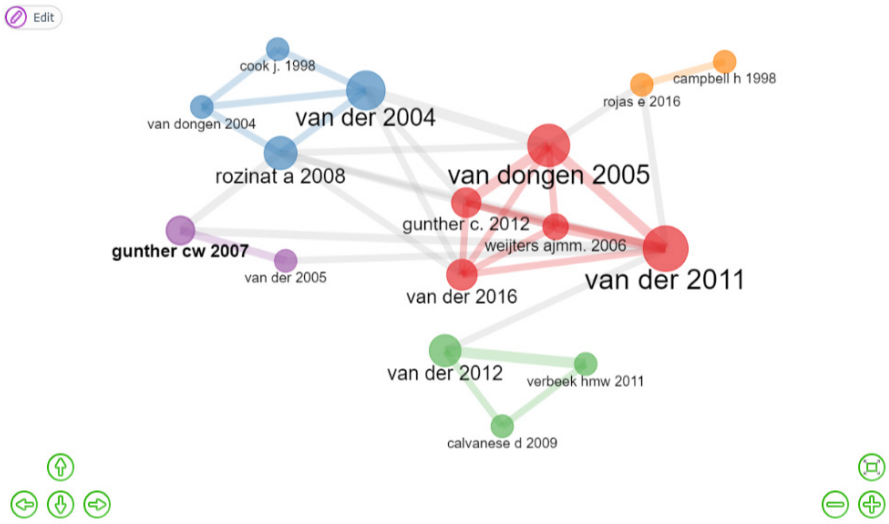


Source: Author's elaboration, 2022.

Intellectual Structure

An analysis of **FIGURE 12** shows that Wil van der Aalst is the most frequently co-cited author, establishing connections across multiple studies and serving as a central reference point within the research field.

FIGURE 12 –Intellectual Structure: Co-citation Network

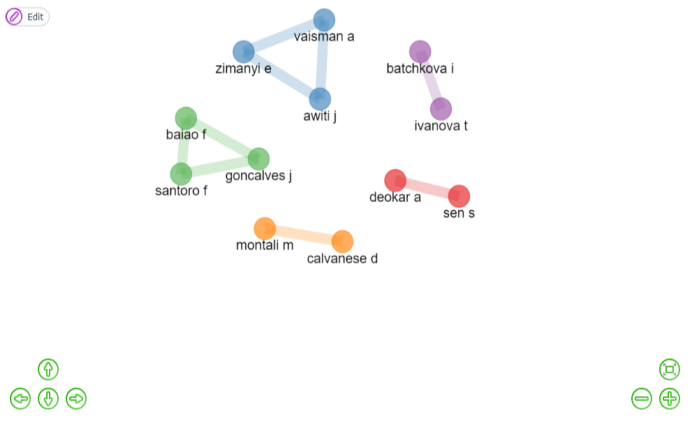


Source: Author's elaboration, 2022.

Social Structure

The social structure illustrates how authors and institutions are interconnected within a specific research domain, typically represented through co-authorship networks (**FIGURE 13**).

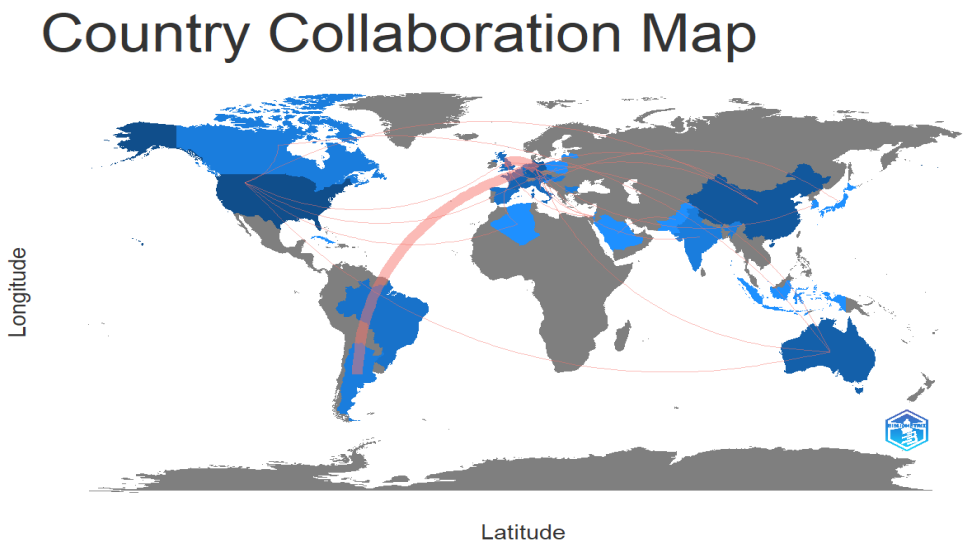
FIGURA 13 – Co-authorship Network



Source: Author's elaboration, 2022.

In terms of international collaboration, the strongest partnerships occur between Belgium and Argentina, and between Germany and the United Kingdom, with several additional, less prominent collaborations also evident (**Figure 14**).

Figura 14 – Country Collaboration Map



Source: Author's elaboration, 2022.

FINAL CONSIDERATIONS

This study successfully achieved its objective by addressing the research question: *What is the most commonly used technique for knowledge extraction based on business process modeling?* The findings indicate that **process mining** is the most widely applied technique for extracting organizational knowledge grounded in process modeling.

The data retrieved from the *Scopus* and *WoS* databases were sufficient to employ the *Biblioshiny* package for *Bibliometrix*, which enabled the identification of the most prolific authors and journals on the topic, the most cited articles, and the evolution of specific keywords over time. Additionally, the methods and techniques used in the ten most relevant studies were examined, demonstrating the application of technological tools for knowledge extraction across organizations and services in different sectors.

The adoption of process mining has grown substantially over the past nine years of the analyzed period. Although the term first appeared in this context in 2005 and remained relatively stable until 2013, its use increased sharply thereafter, reaching 15 occurrences by the end of 2021.

This bibliometric analysis identified significant relationships among studies and revealed trends in the literature, offering a foundation for future systematic reviews on knowledge extraction in organizations and services grounded in business process modeling. The retrieval of only 70 publications from Scopus and 40 from WoS—with 20 duplicates and no time restrictions—suggests that the number of studies on this topic remains limited. This

finding indicates that the field is still emerging and presents considerable potential for further exploration and development.

For future research, it would be valuable to incorporate additional bibliometric indicators and expand the analysis to include other databases, such as *IEEE Xplore* and *Google Scholar*, to produce a more comprehensive quantitative overview. Furthermore, in-depth qualitative analyses of the most cited publications could offer richer insights and strengthen understanding of the theoretical and methodological evolution of knowledge extraction based on business process modeling.

BIBLIOGAPHY

ABPMP INTERNACIONAL. **Guia para o gerenciamento de processos de negócios, corpo comum de conhecimento**: ABPMP BPM CBOK. Brasília: ABPMP Brasil, 2020.

ALHARBI, G. L.; ALOUD, M. E. The effects of knowledge management processes on service sector performance: evidence from Saudi Arabia. **Humanities and Social Sciences Communications**, [*s. l.*], v. 11, n. 1, 378, p. 1-19, 2024. DOI 10.1057/s41599-024-02876-y.

ARIA, M.; CUCCURULLO, C. Bibliometrix: An R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, [*s. l.*], v. 11, n. 4, p. 959–975, 2017. DOI 10.1016/j.joi.2017.08.007.

BEHESHTIA, A.; BENATALLAH, B.; NEZHAD, H. R. M.; YANG, J.; CASATI, F.; ZHANG, X.; SHENG, Q. Z.; DUSTDAR, S.; XUE, S. ProcessGPT: transforming business process management with generative artificial intelligence. *In*: Proceedings of the 2023 IEEE International Conference on Web Services, 2023, Chicago. **Proceedings** [...]. Chicago: ICWS, 2023. p. 731–739. DOI 10.48550/arXiv.2306.01771.

BENCSIK, A. The sixth generation of knowledge management – the headway of artificial intelligence. **Journal of International Studies**, [*s. l.*], v. 14, n. 2, p. 84-101, 2021. DOI 10.14254/2071-8330.2021/14-2/6.

CALVANESE, D.; MONTALI, M.; SYAMSIYAH, A.; AALST, W. M. P. V. D. Ontology-driven extraction of event logs from relational databases. *In*: REICHERT, M.; REIJERS, H. A. (ed.). **Business process management workshops**: BPM 2015, 13th International Workshops, Innsbruck, Austria, August 31–September 3, 2015, Revised Papers. [*S. l.*]: Springer International Publishing, 2016. v. 256. p. 140-153. (Lecture Notes in Business Information Processing [LNBIP] series).

CLARIVATE. **KeyWords plus generation, creation and changes**. c2025. Available in: https://support.clarivate.com/ScientificandAcademicResearch/s/article/KeyWords-Plus-generation-creation-and-changes?language=en_US. 2022. Accessed at: 3 Feb. 2022.

DESAI, N.; VERAS, L.; GOSAIN, A. Using Bradford's law of scattering to identify the core journals of pediatric surgery. **Journal of Surgical Research**, [*s. l.*], v. 229, p. 90–95, Sept. 2018. DOI 10.1016/j.jss.2018.03.062.

DONTHU, N.; KUMAR, S.; MUKHERJEE, D.; PANDEY, N.; LIM, W. M. How to conduct a bibliometric analysis: an overview and guidelines. **Journal of Business Research**, [*s. l.*], v. 133, p. 285–296, 2021.

DUSTDAR, S.; HOFFMANN, T.; AALST, W. V. D. Mining of ad-hoc business processes with TeamLog. **Data and Knowledge Engineering**, [*s. l.*], v. 55, n. 2, p. 129-158, 2005. ISSN 0169-023X.

GARFIELD, E.; SHER, I. H. Key words plus [TM]-algorithmic derivative indexing. **Journal of the American Society for Information Science**, [*s. l.*], v. 44, n. 5, p. 298, June 1993.

GEIERHOS, M.; SCHULZE, S.; SIMON BÄUMER, F. What did you mean? - facing the challenges of user-generated software requirements. *In*: Proceedings of the International Conference on Agents and Artificial Intelligence, 2015, Lisbon. **Proceedings** [...]. Lisbon: SciTePress, 20215. v. 2. p. 277-283. DOI 10.5220/0005346002770283.

GUO, L.; SHARMA, R.; YIN, L.; LU, R.; RONG, K. Automated competitor analysis using big data analytics: evidence from the fitness mobile app business. **Business Process Management Journal**, [*s. l.*], v. 23, n. 3, p. 735-762, 2017. DOI 10.1108/BPMJ-05-2015-0065.

HARMON, P. **Business process change**: a business process management guide for managers and process professionals. 4. ed. Cambridge: Morgan Kaufmann, 2019.

HIRSCH, J. E. An index to quantify an individual's scientific research output. **Proceedings of the National academy of Sciences**, [*s. l.*], v. 102, n. 46, p. 16569-16572, 2005. DOI 10.1073/pnas.0507655102.

JESTON, J. **Business process management**: practical guidelines to successful implementations. 4. ed. New York: Routledge, 2018.

KAMPIK, T.; WARMUTH, C.; REBMANN, A.; AGAM, R.; EGGER, L. N. P.; GERBER, A.; HOFART, J.; KOLK, J.; HERZIG, P.; DECKER, G.; DER, H. V.; POLYVYANYY, A.; RINDERLEMA, S.; WEBER, I.; WEIDLICH, M. Large process models: a vision for business process management in the age of generative AI. **KI-Künstliche Intelligenz**, [*s. l.*], p. 1-15, 2024. DOI 10.1007/s13218-024-00863-8.

KHATIB, S. F.A.; ABDULLAH, D. F.; HENDRAWATY, E.; ELAMER, A. A. A bibliometric analysis of cash holdings literature: current status, development, and agenda for future research. **Management Review Quarterly**, [*s. l.*], v. 72, n. 3, p. 707-744, 2021. DOI 10.1007/s11301-021-00213-0.

LI, J.; WANG, H. J.; Bai, X. An intelligent approach to data extraction and task identification for process mining. **Information Systems Frontiers**, [*s. l.*], v. 17, n. 6, p. 1195–1208, 2015. DOI 10.1007/s10796-015-9564-3.

LIZANO-MORA, H.; PALOS-SÁNCHEZ, P. R.; AGUAYO-CAMACHO, M. The evolution of business process management: a bibliometric analysis. **IEEE Access**, [*s. l.*], v. 9, p. 51088-51105, 2021.

MIGUEL, M. C.; GERLIN, M. N. M.; DA COSTA, R. P. F. A vigilância tecnológica na base SCOPUS: métricas da produção científica, relações na bibliometria. **ETD-Educação Temática Digital**, [*s. l.*], v. 26, p. 1-23, e024016, 2024. DOI 10.20396/etd.v26i00.8670538.

MIRAFZAL, M.; WADHERA, P.; STAL-LE CARDINAL, J. An exploration of knowledge management activities in multidisciplinary service design organizations. *In*: Proceedings of the International Conference on Engineering Design (ICED23), 23., 2023, Bordeaux. **Proceedings** [...]. Bordeaux: Cambridge University Press, 2023. v. 3. p. 525-534. DOI 10.1017/pds.2023.53.

MOURA, L. K. B.; MESQUITA, R. F.; MOBIN, M.; MATOS, F. T. C.; MONTE, T. L.; LAGO E. C.; FALCÃO C. A. M.; FERRAZ, M. Â. A. L.; SANTOS, C. T.; SOUSA, L. R. M. Uses of bibliometric techniques in public health research. **Iranian Journal of Public Health**, [*s. l.*], v. 46, n. 10, p. 1435 -1436, Oct. 2017.

MUNIZ, E. C. L.; DANDOLINI, G. A.; BIZ, A. A. A customer knowledge management framework. **Journal of Information and Knowledge Management**, [*s. l.*], v. 20, n. 3, 2150028, 2021. DOI 10.1142/S0219649221500283.

OKOYE, K.; TAWIL, A.-R. H.; NAEEM, U.; BASHROUSH, R.; LAMINE, E. A semantic rule-based approach supported by process mining for personalised adaptive learning. *In*: The 5th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2014), 5., 2014, Halifax. **Proceedings** [...]. Canada: Procedia Computer Science, 2014. v. 37. p. 203-210. DOI 10.1016/j.procs.2014.08.031.

RIEHMANN, P.; HANFLER, M.; FROEHLICH, B. Interactive sankey diagrams. *In*: IEEE Symposium on Information Visualization (INFOVIS), 2005, Minneapolis. **Conference** [...]. Minnesota: IEEE, 2005. p. 233-240.

ROOS, M.; MARSHALL, M. S.; GIBSON, A. P.; SCHUEMIE, M.; MEIJ, E.; KATRENKO, S.; HAGE, W. R. V.; KROMMYDAS, K.; ADRIAANS, P. W. Structuring and extracting knowledge for the support of hypothesis generation in molecular biology. **BMC bioinformatics**, [*s. l.*], v. 10, S9, p. 1-12, 2009. Suplemento 10. DOI 10.1186/1471-2105-10-S10-S9.

SANTOS, N.; RADOS, G. J. V. **Fundamentos teóricos da gestão do conhecimento**. Florianópolis: Pandion, 2020. E-book. ISBN: 978-65-86527-01-8. Available in: https://www.editorapandion.com/fundamentos-teoricos-de-gestao-do-conhecimento. Accessed at: 15 Dec. 2021.

SCHÖNIG, S.; CABANILLAS, C.; JABLONSKI, S.; MENDLING, J. Mining the organisational perspective in agile business processes. *In*: GAALOUL, K.; SCHMIDT, R.; NURCAN, S.; GUERREIRO, S.; MA, Q. (ed.). **Enterprise, Business-Process and Information Systems Modeling**. BPMDS EMMSAD 2015. Lecture Notes in Business Information Processing. Stockholm: Springer, 2015. v. 214, p. 37-52. DOI doi.org/10.1007/978-3-319-19237-6_3.

TAKEUCHI, H.; NONAKA, I. **Gestão do Conhecimento**. Porto Alegre: Bookman, 2008. *E-book*.

VAN DER AALST, W. M. P. Process mining: a 360 degree overview. In: van der AALST, W.M.P., Carmona, J. (ed.). **Process Mining Handbook**. [*S. l.*]: Springer International Publishing, 2022. p. 3-34. v. 448.

WANG, Y.; CARON, F.; VANTHIENEN, J.; HUANG, L.; GUO, Y. Acquiring logistics process intelligence: methodology and an application for a chinese bulk port. **Expert Systems with Applications**, [*s. l.*], v. 41, n. 1, p. 195-209, Jan. 2014. DOI 10.1016/j.eswa.2013.07.021.

WEIDLICH, M.; ZIEKOW, H.; GAL, A.; MENDLING, J.; WESKE, M. Optimizing event pattern matching using business process models. **IEEE Transactions on Knowledge and Data Engineering**, [*s. l.*], v. 26, n. 11, p.

2759-2773, Nov. 2014. DOI 10.1109/TKDE.2014.2302306.

ZUPIC, I.; CATER, T. Bibliometric methods in management and organization. **Organizational Research Methods**, [*s. l.*], v. 18, n. 3, p. 429-472, 2015. DOI 10.1177/1094428114562629.