 doi:10.20944/preprints202210.0044.v1

Identification of promising research issues in the Digital Industry topic by analyzing Scopus bibliometric data for 2018–2022

B.N. Chigarev
Oil and Gas Research Institute, Russian Academy of Sciences, Moscow, Russia
E-mail: bchigarev@ipng.ru; ORCID: 0000-0001-9903-2800

Annotation

The article is devoted to identifying promising research issues on the topic of Digital Industry based on the analysis of bibliometric data from Scopus platform for 2018–2022. The definition of promising issues derived from the titles, additionally abstracts, of highly cited works by authors with high publication activity. Authors' Scopus profiles were used to determine their affiliations, which are important for in-depth analysis of selected promising Digital Industry tasks. Metadata filtering to identify highly cited papers on particular issues was performed by keywords, Subject Areas, affiliations with countries and sponsoring organizations. Additionally, bibliometric metadata were clustered based on the co-occurrence of terms using VOSviewer. Within these clusters, articles of interest were identified for the formulation of promising research issues on the topic of Digital Industry, which will be subjected to in-depth analysis in separate papers.

Keywords: digital industry, promising research issues, bibliometric metadata, high-cited publications, Scopus, VOSviewer.

Objectives

This publication launches a long-term study that utilizes bibliometric data from scholarly publications to analyze the promising challenges of digitalization of industry.

The aims of this publication are:

- provide a brief overview of researches on digital industry subjects over the last 5 years, using Scopus data
- identify some research issues that require more detailed analysis in the future, e.g., digital manufacturing, Industry 4.0, machine learning in the oil industry
- to suggest authors and their articles to start a detailed analysis of the selected issues

Research Study Motivation

Identifying promising research objectives involves reviewing the broad context of existing work on the topic under study, especially that which has achieved international recognition.

Such a possibility, in particular, is provided by abstract databases which allow exporting large amounts of bibliometric metadata for their further analysis. The present paper uses data from Scopus platform, which indexes highly ranked publications and is available to Russian researchers.

Further, I intend to use open access platforms, including dimensions.ai, lens.org, semanticscholar.org, ieeexplore.ieee.org, and openalex.org, that index a broader range of scholarly publications, such as preprints and academic journals not indexed by Web of Science and Scopus, as well as conference proceedings.

Remark

Tables and graphs based on bibliometric data are mostly self-explanatory, so it is not very useful to fill an article with a detailed discussion of what they show and what they mean. Brief author's notes are more rational, allowing the subject-matter expert to decide for himself what is important to him in the tables, it is useful to give examples of the most cited publications on which the data presented in the tables are based, give information about their authors. The ultimate carriers of knowledge are the authors of publications, so it is advisable to focus on researchers involved in promising issues and their affiliations.

More comprehensive high-resolution tables and graphs can be made available to readers, for example by placing them on figshare.com or by making the data available for viewing on specialized services (app.vosviewer.com).

In my opinion, the structure of this article can be based on the MDPI DATA journal format.

In terms of content, the format of a thematic analysis article is a good choice as it is based on research queries to abstract databases and identifying promising themes based on subjective assessments. It is advisable to limit the thematic analysis to a semantic approach, as it is based on analysing the raw texts rather than identifying their underlying meanings.[1]

Introduction

The Digital Industry theme has the following signs of a new, in-demand topic:

- The ratio of the number of publications in 2021 and 2018 for the query TITLE-ABS-KEY (digital AND industry) AND PUBYEAR > 2017 (the year 2022 has not yet ended and not all publications are indexed): 7845/3558 = 2.2, while for more general queries the increase in the number of publications is less pronounced, for TITLE-ABS-KEY (industry) over the same time period 125893/88894 = 1.4 and for TITLE-ABS-KEY (economy) 48671/31638 = 1.5

- New topics are characterized by a higher ratio of document types Conference Paper/ Article in the abstract database; for the topic digital AND industry, it is 12168/11618=1.047, while for industry it is 129299/298064 = 0.434 and for economy it is 30347/121501 = 0.25.

The publications closest to the content of this paper meet the queries:

TITLE-ABS-KEY (digital AND industry) AND PUBYEAR > 2017 AND TITLE (“thematic analysis”) AND (LIMIT-TO (LANGUAGE, “English”)) for which as of 13-07-2022 there are 4 publications in Scopus and only three of them have DOIs, the fourth is not an article but conference proceedings, let's analyze only the articles.

TITLE-ABS-KEY (digital AND industry) AND PUBYEAR > 2017 AND TITLE (“bibliometric analysis”) AND (LIMIT-TO (LANGUAGE, “English”)) with 44 results. Consider the three most cited publications.

Articles having the terms "thematic analysis" and "bibliometric analysis" in their titles may be similar in content. The difference lies rather in the scope of application - "bibliometric analysis" is more often used as a quantitative analysis of publication metadata and "thematic analysis" as a qualitative analysis of the texts themselves. But once the texts have been annotated in "thematic analysis", quantitative methods widely used in "bibliometric analysis", especially those based on the co-occurrence of terms, can be applied. The clustering of terms or content-related documents, the use of controlled vocabularies (which can be used to annotate texts as well) converge the basic approaches used in both methods of analysis.

Of course, citation is not about "thematic analysis", it is about selecting data for analysis. The research results depend on the questions we seek to answer using the content of the databases of scientific publication abstracts and how we choose the filtering criteria and pre-process the data. It is more important to find acceptable answers to the questions posed, which are subjective in nature, than simply to obtain quantitative estimates. For example, when collecting data on a topic of interest, it is more useful to select terms and filters to refine queries than to simply quantify the occurrence of terms. The objective lies beyond the analysis being conducted.

Among the three papers containing the term "thematic analysis" in their title, the most cited

[1] https://gradcoach.com/what-is-thematic-analysis

(24) paper [1] deals with the coronavirus pandemic, its relation to the digital industry is very indirect – through the digitalization of big data processing. This is a typical automatic classification problem due to the use of Index keywords.

The article [2] is the opposite of the first - it is cited twice, published in 2022, and its title can be interpreted as the relevant issue of Digital Industry: “Toward digital construction supply chain-based Industry 4.0 solutions: scientometric-thematic analysis”. Supply chain-based Industry 4.0 solutions are an obvious issue, just think of the challenges in logistics during the pandemic and the sanctions; the digitalization of logistics is necessary to quickly adapt solutions in a fast-changing environment.

Another paper, published in 2022 with the term “thematic analysis” in the title [3], looks at Augmented Reality in Design and Construction, and could also serve as a basis for more detailed analysis in the broader context of the digital industry, not just construction but also the oil and gas industry.

Bibliometric studies are not only more frequent (44 vs. 4), but are also significantly more often cited. This is largely interrelated; the more researchers are involved in a topic, the more likely that a good paper will be cited. For mediocre papers, this statement can hardly be made. The journal of publication has a significant impact on the citation rate. Engineering Applications of Artificial Intelligence published 357 papers in 2021 and is Q1 in the SCImago Journal Rank and Smart and Sustainable Built Environment - 78 papers in 2021 and is also in Q1. More authors publish in the first journal, they read the journal in which they publish and are more likely to cite an article published in that journal.

The most cited work corresponding to the above query for bibliometrics [4] (220 citations) also deals with research on Industry 4.0. This publication is a classic bibliometric work combined with a thematic review and can be useful as a starting point for more detailed research on the individual tasks listed therein.

The article [5] provides a comprehensive review of BIM-GIS integration for the analysis and practical application of data integration technology, building energy management and urban management. Research trends in the application of BIM-GIS integration are discussed. It is concluded that comprehensive building energy management information should be digitized and quantified for integration and application in urban environment. The value of this paper, in my opinion, is that it considers the real challenge of digitization of urban and energy management of buildings.

The authors [6] address the challenge of understanding digital transformation (DT) trends by using a machine learning (ML) approach to thematic modeling and analysis of advanced research and development in the field of DT technologies. A methodology for compiling a systematic review based on the collected literature is developed. Six dominant themes are identified: smart factory, sustainability and product-service systems, digital transformation in construction, digital infrastructure transformation, technology-driven digital transformation, and business model-driven digital transformation. Crucially for us, the authors demonstrate the usefulness of thematic modeling in identifying the promising issues of industry digitalization.

Thus, a brief review of published works shows the relevance of applying thematic analysis of bibliometric data for the analysis of promising issues of digital industry.

Materials and methods

The bibliometric data used in this article were obtained from the Scopus abstract database by querying: TITLE-ABS-KEY (digital AND industry) AND PUBYEAR > 2017 AND (LIMIT-TO (SRCTYPE, “j”)) AND (LIMIT-TO (LANGUAGE, “English”)), here j stands for journal article, only publications in English were used, which simplifies textual analysis. The data are up-to-date as of 17 June 2022. A total of 12,452 articles were indexed in Scopus for this query. In comparison, the TITLE-ABS-KEY (digital AND industry) query indexed 59,428 publications in Scopus, of which 22,777 were in the form of journal articles in English.

Using only journal articles metadata not only makes the sample more homogeneous, but practice also shows that the database fields for such publications are the most populated.

For other purposes, e.g., to determine the growth of interest in a topic, it is useful to compare the metadata of conference proceedings, articles, and chapters. In conference proceedings, new topics often appear earlier than in articles and book chapters.

Since Scopus allows only 2,000 records to be exported at a time, for 2019–2022 the 2,000 most cited records were exported, and for 2018 all 1,432 records were exported. Thus, a total of, 9432 bibliometric entries were collected for analysis.

Exported data in summary tables, called Scopus_exported_refine_values files, were used to analyze the dynamics of a series of indicators from year to year. The tables in the first part of the Results section were obtained using the SQL Inner Join statement for the annual tables.

We are interested in the common view on publication data for the years 2018–2022, so the Inner Join statement has been used. If there were a different goal, for example, the appearance of new authors or topics, it would be more rational to use the Right Join statement, and the overall picture, respectively, could be obtained with the Full Join statement.

In the second part of the Results section, I used the findings obtained by constructing networks based on term co-occurrence (co-authors can also be treated as terms) and further clustering the data using the Leiden algorithm of the VOSviewer software [7]. In my opinion, it is most convenient to take data from exported MAP and NET files, which can be handled as tsv tables.

There is a link to the json format files exported from VOSviewer at the end of the article, which allows browsing them in detail on the https://app.vosviewer.com/ service. Presenting such results as pictures in the article is impractical because, unlike the app viewer, they cannot be scaled and individual connections between terms cannot be examined.

Materials and methods are limited to the main aim of the article: “to provide a brief overview of research on digital industry topics”, a more detailed treatment of individual issues on digital industry topics is intended to be given in separate publications.

Research results

In the first part of the analysis of bibliometric data retrieved from Scopus on "Digital Industry" produced year-by-year tables for key indicators characterizing publications, such as Subject Areas, Keywords, Source Titles, Author Affiliations with Countries, Author Affiliations with Institutions, и Funding Sponsors. These data allow assessing the basic trends in research areas and the resources involved in sustaining these trends. In the second part of this section, the data from the tables are used to identify Digital Industry Issues, which are worth exploring in more detail in separate studies.

Table 1 shows the top Subject Areas for 2018–2022, the full tables are available on Figshare at the link:

Table 1. Top Subject Areas by years. Retrieved by inner join lists of Subject Areas terms for each year.

SUBJECT AREA	N 2018	N 2019	N 2020	N 2021	N 2022
Engineering	521	840	1098	1444	819
Computer Science	356	607	791	1169	645
Business, Management and Accounting	300	455	610	882	533

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2022 doi:10.20944/preprints202210.0044.v1

Social Sciences	354	538	636	904	477
Environmental Science	175	222	311	450	239
Energy	111	155	284	411	234
Materials Science	118	169	326	357	173
Mathematics	63	84	148	226	158
Physics and Astronomy	73	99	226	262	135
Decision Sciences	67	86	180	250	134
Economics, Econometrics and Finance	99	146	183	251	133
Medicine	106	126	144	219	117
Chemical Engineering	95	87	144	209	115
Chemistry	61	88	127	160	88
Agricultural and Biological Sciences	57	94	109	155	87
Biochemistry, Genetics and Molecular Biology	72	72	118	162	83
Arts and Humanities	96	127	131	174	75
Psychology	24	40	43	94	68
Earth and Planetary Sciences	37	57	113	136	65
Pharmacology, Toxicology and Pharmaceutics	28	33	65	54	33
Health Professions	7	22	24	30	20
Neuroscience	7	7	5	14	19
Immunology and Microbiology	13	11	19	29	17
Multidisciplinary	13	17	30	52	10
Nursing	9	11	8	13	10
Veterinary	5	10	13	14	10
Dentistry	3	2	5	4	3

Subject Areas: Engineering, Computer Science and Business, Management and Accounting are the expected topics for Digital Industry, of particular interest for deeper study is Social Sciences. For this, the additional filter in the Scopus query can be given as: AND (LIMIT-TO (SUBJAREA, "SOCI")).

The social aspects of the Digital Industry, for example, are presented in the article [8] cited 503 times in Scopus on 06 July 2022, one of it' s questions is: How do digital platforms affect everyday life? That is, not about business, but about current life.

A more in-depth analysis of the importance of the **social dimension for the Digital Industry** can start by reviewing the profiles of the authors of the publication.

Profiles of the authors of this paper:

Reuver, Mark De; Faculteit Techniek, Bestuur en Management, TU Delft, Delft, Netherlands; Scopus ID 35776955900; https://orcid.org/0000-0002-6302-7185; 111 Documents by author; 2142 Citations by 1876 documents; 23 h-index.

Sørensen, Carsten; London School of Economics and Political Science, London, United Kingdom; Scopus ID 7102879062; https://orcid.org/0000-0002-2002-9383; 67 Documents by author; 2833 Citations by 2244 documents; 21 h-index.

Basole, Rahul C.; Accenture, New York, United States; Scopus ID 24329464200; 80 Documents by author; 2437 Citations by 1895 documents; 22 h-index.

The importance of human capital in the Digital Industry topic is outlined in [9], cited by 170 documents.

Profiles of the authors of this paper:

Popkova, Elena G.; Moscow State Institute of International Relations (MGIMO), Moscow, Russian Federation; Scopus ID 55671568200; https://orcid.org/0000-0003-2136-2767; 202 Documents by author; 3565 Citations by 1419 documents; 33 h-index.

Sergi, Bruno S.; Harvard University, Cambridge, United States; Scopus ID 35369323900; https://orcid.org/0000-0002-5050-5651; 170 Documents by author; 1872 Citations by 1213 documents; 23 h-index.

The top keywords (Authors and Index) for 2018-2022 are listed in Table 2.

Table 2. Top Keywords (Authors and Index) by years. Retrieved by inner join lists of Keyword terms for each year.

KEYWORD	N 2018	N 2019	N 2020	N 2021	N 2022
Digital Storage	260	302	435	551	310
Industry 4.0	79	199	373	501	324
Digital Transformation	28	88	158	303	172
Human	128	148	187	276	126
Digital Twin	10	44	133	267	126
Article	117	110	172	228	101
Artificial Intelligence	35	61	120	217	123
Humans	82	102	105	181	86
Internet Of Things	63	107	137	179	147
Digital Technologies	33	51	105	179	88
Blockchain	16	50	74	171	89
Machine Learning	22	46	88	156	96
Decision Making	51	58	117	141	89
Construction Industry	32	46	72	135	84
Digitalization	32	62	88	132	108
Sustainability	31	37	67	129	82
Innovation	43	76	83	124	65
Sustainable Development	23	34	71	117	91
Information Management	28	49	61	110	60
Big Data	58	87	83	105	78
Manufacture	38	57	77	104	63
Life Cycle	18	39	59	88	59
Supply Chains	25	25	49	87	50
3D Printers	26	44	64	86	40
Digitization	25	40	63	85	43
Technology	35	37	55	83	47
Automation	31	35	58	82	62
Review	25	42	60	78	33
Industrial Research	23	30	59	78	74
Manufacturing	14	32	48	78	54

Most of the keywords in the table are clearly related to the Digital Industry topic. However, it is worth clarifying why publications are so often categorized as Digital Storage by the Scopus platform in all the years under consideration.

Using the metadata of 1,877 publications matching the query: TITLE-ABS-KEY (digital AND industry) AND PUBYEAR > 2017 AND KEY ("Digital Storage") AND (LIMIT-TO (SRCTYPE, "j")) AND (LIMIT-TO (LANGUAGE, "English")), the term Digital Storage appears only 4 times in the Title, Abstract and Author keywords, more specifically only in the abstract and in phrases such as using Digital Storage Oscilloscope Meanwhile, in Index Keywords, the term Digital Storage is found in all records. Therefore, the term Digital Storage is strictly a Scopus Index Keywords term and reflects the classification of records in the system, not the occurrence of the term in publication metadata.

The topic that Scopus refers best as "Digital Storage" is reflected in the title of the article [10], "**Digital twin-driven product design, manufacturing and service with big data**", which describes the issue fitting well for a separate study.

The profiles of the authors of this article are used below and listed after Table 8.

The top Source Titles for 2018-2022 are listed in Table 3.

Table 3. Top Source Titles by years. Retrieved by inner join lists of Source Titles for each year.

SOURCE TITLE	N	N	N	N	N

	2018	2019	2020	2021	2022
Sustainability Switzerland	23	23	62	141	72
IEEE Access	20	32	71	72	26
Applied Sciences Switzerland	5	12	38	70	26
Technological Forecasting and Social Change	8	10	14	41	38
Energies	6	20	24	40	25
Journal Of Cleaner Production	11	10	24	28	21
International Journal of Advanced Manufacturing Technology	8	19	13	22	18
Automation In Construction	11	10	15	17	13
Computers In Industry	5	13	12	17	20
International Journal of Production Research	3	4	10	17	19
Materials	3	6	8	15	4
Engineering Construction and Architectural Management	5	6	5	13	8
IEEE Transactions on Industrial Informatics	4	3	10	12	18
International Journal of Computer Integrated Manufacturing	2	6	8	12	5
Plos One	5	3	5	10	4
Advanced Engineering Informatics	3	2	4	10	5
Energy	6	4	7	9	5
Buildings	2	3	3	9	14
Applied Energy	7	13	10	6	6
Multimedia Tools and Applications	3	5	6	6	8
Publishing Research Quarterly	5	8	4	6	4
Computers And Electronics in Agriculture	2	3	7	5	6
International Journal of Advanced Computer Science And Applications	3	3	7	5	3

The top three journals in the table – Sustainability Switzerland, IEEE Access, and Applied Sciences Switzerland published in 2021, respectively: 14,051, 12,850, and 11,975 open access articles. It makes these journals attractive for collecting material on the topic of the Digital Industry. IEEE Access is indexed on the open platform IEEE Xplore[2], and Sustainability Switzerland and Applied Sciences Switzerland on the platform of MDPI Publisher, which has its own effective search engine.

The top Author Affiliations with Countries for 2018–2022 are listed in Table 4.

Table 4. Top Author Affiliations with Countries by years. Retrieved by inner join lists of Author Affiliations with Countries for each year.

COUNTRY	N 2018	N 2019	N 2020	N 2021	N 2022
United States	319	383	459	599	272
China	128	208	350	529	396
United Kingdom	168	231	293	473	233
India	104	198	232	308	209
Germany	88	134	184	275	144
Italy	59	107	174	241	149
Australia	101	117	152	238	145
Spain	57	70	138	199	88
Russian Federation	42	119	107	152	48
France	39	55	88	147	76
Canada	35	76	88	139	76
Brazil	33	45	73	101	74
Malaysia	38	60	62	94	81
South Korea	51	63	57	93	64
Sweden	27	41	57	87	49
Netherlands	35	41	50	84	43
Portugal	16	26	31	77	44

[2] https://ieeexplore.ieee.org/Xplore/home.jsp

Poland	12	24	51	73	45
Saudi Arabia	13	14	39	70	57
Turkey	18	15	33	68	45
Switzerland	30	34	43	66	42
Indonesia	12	55	67	61	30
Finland	27	32	31	56	41
Japan	30	29	34	55	24
Pakistan	6	25	27	53	45
Hong Kong	14	14	28	53	41
Austria	16	31	27	53	25
Denmark	15	24	31	52	41
South Africa	16	18	25	52	36
Taiwan	24	25	52	52	31

The USA was the leading publication destination until 2021, but in 2022 China took the lead. Also noteworthy is the increase in publications from India.

The top Author Affiliations with Institutions for 2018–2022 are listed in Table 5.

Table 5. Top Author Affiliations with Institutions by years. Retrieved by inner join lists of Author Affiliations with Institutions for each year.

AFFILIATION	N 2018	N 2019	N 2020	N 2021	N 2020
Ministry of Education China	8	9	23	41	14
UNSW Sydney	10	6	19	27	10
CNRS Centre National de la Recherche Scientifique	9	17	18	27	16
University College London	4	9	12	25	9
University of Cambridge	12	12	16	24	17
University of Melbourne	9	9	9	24	17
Imperial College London	8	7	13	23	5
National University of Singapore	6	5	7	22	16
Chinese Academy of Sciences	4	8	22	20	12
Universidade de São Paulo	3	8	10	20	12
Queensland University of Technology	4	9	10	20	16
RMIT University	6	8	9	20	15
University of Nottingham	3	5	14	18	8
Deakin University	5	6	13	18	10
Hong Kong Polytechnic University	4	6	14	17	17
Tsinghua University	4	6	11	17	13
Norges Teknisk-Naturvitenskapelige Universitet	4	10	12	16	16
Technical University of Munich	5	5	11	16	7
Curtin University	9	8	7	16	6
The University of Sydney	9	8	15	15	9
King Saud University	5	6	8	15	11
University of Johannesburg	3	5	14	13	20
The University of Manchester	4	11	10	13	12
Monash University	4	12	10	13	11
University of Toronto	4	10	9	13	10
University of Leeds	6	7	8	13	5
Friedrich-Alexander-Universität Erlangen-Nürnberg	4	4	7	13	10
Universiti Kebangsaan Malaysia	3	6	9	12	5
Rheinisch-Westfälische Technische Hochschule Aachen	5	6	8	12	8
University of Technology Sydney	9	7	8	12	8

The affiliation of the institutions of the authors of the publications agrees well with the affiliation by country. In the abstract databases, the fields “Affiliation” and “Foundation” are the least correctly filled in. It is not very useful to analyze the statistics on them. But it is interesting to use the list of affiliated institutions to identify their involvement in specific digital industry tasks.

For example, the article [11], cited 41 times in Scopus, has the following record in the affiliation field "School of Energy Science and Engineering, Harbin Institute of Technology, Harbin, 150001, China; Key Laboratory of Fluid and Power Machinery (Xihua University), **Ministry of Education** Sichuan, Chengdu, 610039, **China**; State Key Laboratory of Hydro-Power Equipment, Harbin Institute of Large Electrical Machinery, Harbin, 150040, China".

The author affiliated with the Chinese Ministry of Education has the following Scopus profile:

Li, Zhenggui; Xihua University, Chengdu, China; Scopus ID 54894723200; 70 Documents by author; 368 Citations by 255 documents; 10 h-index. The Scopus profiles of authors of highly cited publications are usually carefully reviewed, making their use practical.

The metadata of this article is indicative in the context of bibliometric analysis. For example, let us compare the lists of Author and Index Keywords in the record of this publication.

Author Keywords: Bispectrum; Dean Vortex; Frequency characteristic; Pump-turbine; Pumped storage power plant.

Index Keywords: Computational fluid dynamics; Digital storage; Frequency domain analysis; Pumped storage power plants; Shear stress; Turbulence models; Vortex flow; Bispectrum; Dean vortex; Energy storage technologies; Frequency characteristic; Performance characteristics; Pump-turbines; Time- and frequency-domain analysis; Unsteady numerical simulations; Hydraulic turbines; amplitude; energy storage; frequency analysis; numerical method; performance assessment; pump; turbine; turbulence.

The lists above show that Index Keywords significantly expand the set of terms that, according to Scopus, describe the topic of a publication. However, the result is far from consistent, e.g., the term Digital Storage already mentioned earlier – what is its relevance to the topic of the article?

Another problem with the direct use of keywords to identify relevant research issues is the fact that it is difficult to conclude from the lists of terms given why a particular publication is of interest to the Digital Industry topic. However, if we address the title of the article, "Mechanism of high amplitude low frequency fluctuations in a pump-turbine in pump mode", the phrase "Mechanism of high amplitude low frequency fluctuations in a pump" well describes the problem under study. The snippets from the abstract, "Pumped storage technology has become the most important energy storage technology in industry today. The instabilities of pump-turbines as key parts of pumped storage power plants have become critical issues in the development of pump storage technology." clearly explains the "critical issues" and how "unsteady numerical simulations have been carried out", which can be effectively addressed by digitizing the process, from initial information gathering to system simulation.

In my opinion, it is more appropriate to use publication titles to identify topical issues in the digital industry and, if they are not clear enough, to look for explanations in the abstracts. The meaning of the keywords themselves is more relevant in the first stage of research: when querying abstract databases to collect publication metadata.

An article cited in Scopus 12 times [12] illustrates how the title of the paper "Identifying soil management zones in a sugarcane field using proximal sensed electromagnetic induction and gamma-ray spectrometry data" can be reformulated for similar research, since the problem of "**Identifying soil management zones**" can be solved by other technical approaches.

The top Funding Sponsors for 2018–2022 are listed in Table 6.

Table 6. Top Funding Sponsors by years. Retrieved by inner join lists of Funding Sponsors for each year.

 doi:10.20944/preprints202210.0044.v1

FUNDING SPONSOR	N 2018	N 2019	N 2020	N 2021	N 2022
National Natural Science Foundation of China	56	86	163	193	122
European Commission	30	52	85	114	71
Horizon 2020 Framework Programme	18	36	68	69	26
National Key Research and Development Program of China	6	6	38	63	32
European Regional Development Fund	12	15	39	51	17
Engineering and Physical Sciences Research Council	29	39	45	50	18
National Science Foundation	22	26	37	42	16
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior	7	16	21	40	26
Fundamental Research Funds for the Central Universities	9	23	24	35	15
Australian Research Council	13	16	22	35	12
Conselho Nacional de Desenvolvimento Científico e Tecnológico	8	18	30	34	27
National Research Foundation of Korea	16	17	23	34	19
U.S. Department of Energy	12	15	23	32	11
Fundação para a Ciência e a Tecnologia	4	12	13	32	18
UK Research and Innovation	19	21	33	28	19
Ministry of Science and Technology of the People's Republic of China	6	15	25	25	17
Natural Sciences and Engineering Research Council of Canada	5	7	19	25	13
Ministry of Education of the People's Republic of China	11	21	24	23	14
Horizon 2020	6	10	20	23	17
Bundesministerium für Bildung und Forschung	9	8	15	22	11
Russian Foundation for Basic Research	5	16	12	21	4
Ministry of Education	9	8	11	20	3
Deutsche Forschungsgemeinschaft	4	9	15	19	12
Government of Canada	2	14	13	18	8
Ministerio de Economía y Competitividad	7	14	8	18	5
China Postdoctoral Science Foundation	5	7	8	15	9
European Social Fund	2	2	4	15	11
Ministry of Science and Technology, Taiwan	7	7	17	14	8
National Institutes of Health	11	12	7	14	9
Ministry of Finance	4	12	11	12	6

The data in Table 6 can be used to search for projects funded by these donors. For example, Horizon 2020 funds various projects, search GOOLE for projects related to the Digital Industry. Looking at the project “the DAIS project will research and deliver Distributed Artificial Intelligent Systems” with a section “Digital Industry” and find the issue “**Energy management based on a fog computing approach**”. My personal experience is that it is more appropriate to formulate tasks within a topic rather than the topics themselves in order to collect publications in abstract databases. So, there are no publications in Scopus that match the TITLE-ABS-KEY (DIAS AND “digital industry”), but there are 611 documents that match the TITLE-ABS-KEY (energy AND management AND fog AND computing). One of the papers matching the above request [13] and published in 2017 has 2235 citations. The authors' affiliation is also of interest: Hong Kong University of Science and Technology and Hamad Bin Khalifa University, Doha, Qatar.

The following conclusion can be made from the analysis of the tables in this section: the summarized data presented in the tables are of limited interest for identifying current issues for the Digital Industry, but are useful filters for querying abstract databases.

The second part of the study concerns the identification of the authors of highly cited articles and how relevant research issues can be formulated on the basis of the wording of the titles of these articles.

Approaches to solving this problem can be diverse. This article considers only the use of co-occurrence of Authors and Index Keywords in articles and their clustering with the Leiden algorithm, built into the free program VOSviewer [7].

Clustering enables to create groups of coauthors or keywords based on their co-occurrence in metadata.

There are many other methods for forming groups based on the proximity of their elements, but this goes far beyond the scope of one article. You can use clustering of documents by comparing their metadata or use algorithms like Apriori.

Since identifying authors by their last name and first name can be ambiguous, especially in the case of Chinese authors, it is appropriate to use the Author ID.

Authors ID co-occurrence

To maintain the general approach of using VOSviewer to cluster key terms, at this point the Author(s) ID field was renamed to Index Keywords, the data was analyzed and exported, and the field was again renamed to Author(s) ID. Additionally, the following fields were used: Authors, Title, Year, Cited by, and DOI. In the second step, the Index Keywords field itself was used.

Average normalized citation was used to select the authors, and then the most cited publications for these authors were selected. Promising issues for a further, more detailed analysis were formulated according to the publication titles (optionally the abstracts). Brief data on the author's profile from Scopus system were given, allowing to understand the author's affiliation and indicators of his/her publication activity. Both factors are important when gathering information on an issue. Such a choice always carries elements of subjectivity – if not, there will be many choices.

VOSviewer program parameters: total values of Author(s) ID – 32008, occurring more than three times – 787, the minimum number of Author(s) ID in a cluster – 20. Three clusters were obtained with these values, which are discussed below.

Cluster 1

Table 7 shows the 20 Author IDs from the first cluster, sorted by Average normalized citation. The first three Author IDs in bold in the table will be used below.

Table 7. Top Author ID in cluster #1 by Average normalized citation. Occurrences – number of publications by author. Avg. pub. Year – average publication year.

Author ID	cluster	Occurrences	Avg. pub. year	Avg. norm. citations
57192100009	1	3	2019.3333	23.0898
57193673728	1	6	2019	22.5854
12141248300	1	11	2019.6364	15.7347
57195245075	1	6	2020.1667	14.2821
37102014400	1	6	2020.5	12.2811
24921483100	1	3	2020	11.006
57203151995	1	3	2020	11.006
56173081300	1	6	2019.6667	9.1559
57207454205	1	8	2020	7.1859
7801563868	1	3	2020.6667	7.0217
55080150600	1	7	2020.2857	6.8247
57188970257	1	3	2021	5.2968
57219656467	1	4	2020	5.2332
54895137000	1	3	2020.6667	5.044
55101071200	1	3	2021.3333	3.8914
44261803600	1	6	2020.1667	3.7732
36141576300	1	14	2020.8571	3.4892

56948414800	1	3	2021.3333	3.3327
56996337400	1	3	2021.3333	3.3327
57191981777	1	3	2021.3333	3.3327

To get a list of the author's publications by their ID, I used a query to the general table of publications in the form of: select … from … where Author(s) ID like '%57192100009%'

Author Profile from Scopus: Zhang, Meng; Tsinghua University, Beijing, China; Scopus ID 57192100009; 23 Documents by author; 3244 Citations by 2130 documents; 16 h-index.

Next, the most cited articles of the author were taken from the obtained selection.

Table 8 shows the titles and citations of two works by this author. The titles of the articles reflect their subject. The high citation rate, numerous publications, and affiliation with one of the leading universities in the PRC guarantee the promising nature of the issue under study and the possibility of further gathering of publications for a more detailed analysis.

Table 8. Examples of the most cited publications of Author with ID=57192100009 and their co-authors, Authors ID, Title and DOI.

Authors	Author(s) ID	Title. DOI	Cited by
Tao F., Cheng J., Qi Q., Zhang M., Zhang H., Sui F.	12141248300, 57192081753, 57193673728, **57192100009**, 57190839107, 57189511188	Digital twin-driven product design, manufacturing and service with big data. [10]	1059
Tao F., Sui F., Liu A., Qi Q., Zhang M., Song B., Guo Z., Lu S.C.-Y., Nee A.Y.C.	12141248300, 57189511188, 37102014400, 57193673728, **57192100009**, 57201022514, 57201033062, 25933160000, 57195245075	Digital twin-driven product design framework. [14]	322

From the title of the articles, it is clear that "**Product design based on a digital twin**" is a hot topic.

Comparable data for the author with ID 57193673728 are presented in Table 9. The first and third most cited articles of this author are the same as those given above in Table 8 and will not be considered.

Author Profile from Scopus: Qi, Qinglin; Beihang University, Beijing, China; Scopus ID 57193673728; https://orcid.org/0000-0002-3247-0440 29 Documents by author; 4617 Citations by 3083 documents; 18 h-index.

Table 9. Examples of the most cited publications of Author with ID=57193673728 and their co-authors, Authors ID, Title and DOI.

Authors	Author(s) ID	Title. DOI	Cited by
Tao F., Cheng J., Qi Q., Zhang M., Zhang H., Sui F.	12141248300, 57192081753, **57193673728**, 57192100009, 57190839107, 57189511188,	Digital twin-driven product design, manufacturing and service with big data. [10]	1059
Qi Q., Tao F.	57193673728, 12141248300	Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. [15]	550
Tao F., Sui F., Liu A., Qi Q., Zhang M., Song B., Guo Z., Lu S.C.-Y., Nee A.Y.C.	12141248300, 57189511188, 37102014400, **57193673728**, 57192100009, 57201022514, 57201033062, 25933160000 57195245075,	Digital twin-driven product design framework. [14]	322
Tao F., Qi Q., Wang L.,	12141248300, **57193673728**, 55080150600, 57195245075	Digital Twins and Cyber–Physical Systems toward Smart Manufacturing and Industry	271

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2022 doi:10.20944/preprints202210.0044.v1

Nee A.Y.C.		4.0: Correlation and Comparison. [16]	

Digital Twin, Big Data, Cyber-Physical Systems and Smart Manufacturing – another wording of a promising topic.

Fei Tao is the first author of all the above articles, so it is reasonable to cite his profile in Scopus (Beihang University, Beijing, China; Scopus ID 12141248300; https://orcid.org/0000-0002-9020-0633 280 Documents by author; 17177 Citations by 8534 documents; 64 h-index) and use his ID to collect publications.

The promising topic of the first cluster can be summarized as "**Digital Twin for Manufacturing**".

Cluster 2

A similar procedure is performed for the second cluster, where the IDs of the most cited authors are presented in Table 10. Here, as an example, we select two authors with the highest Average normalized citation and one with the largest number of publications (10) and Average normalized citation equal to 6.6938.

Table 10. Top Author ID in cluster #2 by Average normalized citation. Occurrences – number of publications by author. Avg. pub. Year – average publication year.

Author ID	cluster	Occurrences	Avg. pub. year	Avg. norm. citations
55599784700	2	5	2019.6	18.5453
54790849500	2	7	2020	13.7009
15757819200	2	4	2020.25	11.0723
54977014200	2	4	2020.5	7.8203
36864045000	2	8	2020	7.772
8417810000	2	3	2019.6667	7.712
9279823400	2	3	2020.6667	7.3767
26429368200	2	3	2020	7.3301
56238759300	2	10	2020.2	6.6938
57192378881	2	3	2021	6.2634
55757488200	2	10	2020.5	6.0734
24831905500	2	3	2019.6667	5.9432
57201069320	2	3	2020.6667	5.4164
55851943244	2	9	2021.1111	4.9571
56370773200	2	3	2019.6667	4.3227
42262603800	2	4	2021.25	3.3407
55735821600	2	4	2020.5	3.1253
57191022246	2	8	2021	3.1245
57193437808	2	3	2021.6667	3.1008
14833520200	2	3	2020.3333	3.053

The titles of articles by the author with Id 55599784700 are available in Table 11.

Author Profile from Scopus: Ayala, Néstor Fabián; Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; Scopus ID 55599784700; https://orcid.org/0000-0001-8888-9227. 26 Documents by author; 1943 Citations by 1447 documents; 11 h-index.

Table 11. Examples of the most cited publications of Author with ID=55599784700 and their co-authors, Authors ID, Title and DOI.

Authors	Author(s) ID	Title. DOI	Cited by
Dalenogare L.S., Benitez G.B., Ayala N.F., Frank A.G.	57203842897, 57197831648, **55599784700**, 54790849500,	The expected contribution of Industry 4.0 technologies for industrial performance. [17]	568
Frank A.G., Dalenogare L.S., Ayala N.F.	54790849500, 57203842897, **55599784700**,	Industry 4.0 technologies: Implementation patterns in manufacturing companies. [18]	707

The promising topic from these titles can be summed up as follows: **Industry 4.0**

technologies for industrial performance.

Frank A.G. (ID 54790849500) co-authored the articles presented in Table 11, so Table 12 gives his next two most-cited articles.

Author Profile from Scopus: Frank, Alejandro Germán; Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; Scopus ID 54790849500; https://orcid.org/0000-0001-5041-6467; 63 Documents by author; 2955 Citations by 2228 documents; 23 h-index.

Table 12. Examples of the most cited publications of Author with ID=54790849500 and their co-authors, Authors ID, Title and DOI.

Authors	Author(s) ID	Title. DOI	Cited by
Frank A.G., Mendes G.H.S., Ayala N.F., Ghezzi A.	**54790849500**, 56785594200, 55599784700, 24831905500,	Servitization and Industry 4.0 convergence in the digital transformation of product firms: A business model innovation perspective. [19]	271
Benitez G.B., Ayala N.F., Frank A.G.	57197831648, 55599784700, **54790849500**,	Industry 4.0 innovation ecosystems: An evolutionary perspective on value cocreation. [20]	112

The promising topic from these titles can be summed up as follows: **Industry 4.0 and innovations**.

The author with ID 56238759300 has one of the biggest number of publications, but a medium Average normalized citation, his profile in Scopus: Gunasekaran, Angappa; California State University, Los Angeles, Los Angeles, United States; Scopus ID 56238759300. 571 Documents by author; 33042 Citations by 22767 documents; 96 h-index.

His large total number of 571 papers and his affiliation with a strong laboratory at a prominent university are worth noting. Two of his papers are presented in table 13.

Table 13. Examples of the most cited publications of Author with ID=56238759300 and their co-authors, Authors ID, Title and DOI.

Authors	Author(s) ID	Title. DOI	Cited by
Kamble S.S., Gunasekaran A., Gawankar S.A.	36864045000, 56238759300, 55846962200,	Sustainable Industry 4.0 framework: A systematic literature review identifying the current trends and future perspectives. [21]	464
Kamble S.S., Gunasekaran A., Sharma R.	36864045000, 56238759300, 57196545771,	Analysis of the driving and dependence power of barriers to adopt industry 4.0 in Indian manufacturing industry. [22]	229

The promising topic from these titles: Sustainable Industry 4.0.

The promising topic for Cluster 2: **Sustainable, Innovative, Technology-efficient Industry 4.0**.

Cluster 3

The Average normalized citation of the publications of the authors whose list is presented in Table 14 is lower than in the first two clusters. Understanding the underlying reasons for the lower citation rate is also crucial for choosing a promising research topic.

Table 14. Top Author ID in cluster #3 by Average normalized citation. Occurrences – number of publications by author. Avg. pub. Year – average publication year.

Author ID	cluster	Occurrences	Avg. pub. year	Avg. norm. citations
56687825200	3	3	2020.3333	4.8081
56349650400	3	3	2020	4.7602
57195542948	3	3	2020.3333	4.7487
57021768700	3	3	2021.3333	4.6685
35239818900	3	7	2021.1429	4.6676
6508058214	3	3	2021.3333	4.5874

55348625200	3	4	2020	4.3028
35310169200	3	4	2021	4.2821
21743792400	3	6	2019.5	4.2438
35427952100	3	3	2020	3.9212
55808097300	3	3	2021.6667	3.8566
14037119900	3	4	2019.5	3.8353
57194166860	3	6	2019.6667	3.7957
50262828700	3	4	2021.5	3.7849
26422351200	3	3	2020	3.7542
15726654000	3	4	2020.25	3.7464
56595145400	3	4	2020.25	3.7464
57224222999	3	4	2020.5	3.6195
55911169300	3	6	2021.1667	3.5298
55960201200	3	4	2020.5	3.0689

Two authors with the highest Average normalized citation and one with 7 publications are selected from this table.

Author (56687825200) Profile from Scopus: Rosa, Paolo; Politecnico di Milano, Milan, Italy; Scopus ID 56687825200; https://orcid.org/0000-0003-3957-707X. 54 Documents by author; 2046 Citations by 1651 documents; 23 h-index.

Both articles of the author in the table 15 are of reviewing nature.

Table 15. Examples of the most cited publications of Author with ID=56687825200 and their co-authors, Authors ID, Title, DOI and Citation.

Authors	Author(s) ID	Title. DOI	Cited by
Rosa P., Sassanelli C., Urbinati A., Chiaroni D., Terzi S.	**56687825200**, 56595145400, 57194166860, 14037119900, 15726654000	Assessing relations between Circular Economy and Industry 4.0: a systematic literature review. [23]	162
Rocca R., Rosa P., Sassanelli C., Fumagalli L., Terzi S.	57192381259, **56687825200**, 56595145400, 55348625200, 15726654000	Integrating virtual reality and digital twin in circular economy practices: A laboratory application case. [24]	47

The promising topic: **Circular Economy and Industry 4.0**. The Circular Economy is a separate topic for analysis. A more targeted issue is using the capabilities of Industry 4.0 to improve the efficiency of Circular Economy.

Two papers by the author (56349650400) with profiles from Scopus: Negri, Elisa; Politecnico di Milano, Milan, Italy; Scopus ID 56349650400; https://orcid.org/0000-0002-0006-7497. 40 Documents by author; 1521 Citations by 1197 documents; 18 h-index, are given in Table 16.

Table 16. Examples of the most cited publications of Author with ID=56349650400 and their co-authors, Authors ID, Title and DOI.

Authors	Author(s) ID	Title. DOI	Cited by
Cimino C., Negri E., Fumagalli L.	57210824444, **56349650400**, 55348625200	Review of digital twin applications in manufacturing. [25]	174
Negri E., Berardi S., Fumagalli L., Macchi M.	**56349650400**, 57216900878, 55348625200, 7004929933	MES-integrated digital twin frameworks. [26]	37

The promising topic: **Digital twin applications in manufacturing** – the task overlaps with the topic of cluster 1, the clustering is based on co-authorship, not proximity of the texts of publications, so the result is affected by the affiliation of the authors. So, the authors with ID 56349650400 and ID 56687825200 are both from Milan, Italy.

Two papers by the author (35239818900) with profiles from Scopus: Dwivedi, Yogesh K.; Swansea University, Swansea, United Kingdom; Scopus ID 35239818900; https://orcid.org/0000-0002-5547-9990. 528 Documents by author; 20159 Citations by 12107 documents; 76 h-index, are given in Table 17.

Table 17. Examples of the most cited publications of Author with ID=35239818900 and their co-authors, Authors ID, Title and DOI.

Authors	Author(s) ID	Title. DOI	Cited by
Ali O., Ally M., Clutterbuck, Dwivedi Y.	56498722200, 8925161400, 57218221904, 35239818900	The state of play of blockchain technology in the financial services sector: A systematic literature review. [27]	65
Bag S., Pretorius J.H.C., Gupta S., Dwivedi Y.K.	55757488200, 8417810000, 55851943244, 35239818900	Role of institutional pressures and resources in the adoption of big data analytics powered artificial intelligence, sustainable manufacturing practices and circular economy capabilities. [28]	109

The Circular Economy topic brings the publications in Tables 15 and 17 into closer alignment, while the publications in the first table focus on the more general issues of Industry 4.0, the publications in the second address more specific ones: blockchain technology, big data, artificial intelligence, and sustainable manufacturing practices.

Remark: Clustering based on co-authorship strongly depends on the authors' affiliation, so the first cluster may be related to China, the second to Brazil, and the third to Europe. It is very important to understand which terms prevail in the publications of a particular region/state, as this makes it easier to collect literature revealing the research and economic priorities of a particular region. The meaning of the content of articles is best reflected in their titles and abstracts; keywords are more needed when making queries to abstract databases.

When formulating promising research tasks, it is advisable to use two or three key terms, rather than one, for example, Circular Economy – a very broad topic, while big data analytics + manufacturing practices + Circular Economy define a particular issue for which one can search for examples of implementations and projects, and this is of most interest to specialized professionals.

Countries co-occurrence

Clustering based on co-occurrence of affiliations with countries for co-authors of publications helps to understand with which countries cooperation is developed and with which it is weak, but there is potential to increase it.

The scope of publications by country varies greatly, and the citation rate can be strongly influenced by publication activity in the native language, so at this stage of the analysis we sort simply by the number of publications.

Table 18 shows the list of countries with the highest number of publications in the first cluster, formed on the basis of affiliations co-occurrence of co-authors by country.

Cluster 1

Table 18. Top Countries by number of publications (Occurrences) in cluster #1

Country	cluster	Occurrences	Avg. pub. year	Avg. norm. citations
india	1	767	2020.2073	1.1072
france	1	334	2020.3174	1.9847
canada	1	334	2020.2425	1.2743

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2022 doi:10.20944/preprints202210.0044.v1

russian federation	1	292	2019.7877	0.7267
malaysia	1	264	2020.2765	0.8247
saudi arabia	1	160	2020.725	1.2119
turkey	1	139	2020.4964	1.3329
united arab emirates	1	98	2020.3571	1.4716
egypt	1	55	2020.4364	1.1889
kazakhstan	1	37	2020.4595	0.4417

Despite the low level of normalized citations, both Russia and Kazakhstan have good publication activity in English, despite the fact that authors from these countries actively publish in Russian.

Multiple sampling options out of numerous bibliometric data can be justified. In this article, the emphasis is placed on identifying the authors most relevant to a particular topic, regardless of the type of clustering: by country, index keywords, or clustering by co-authorship.

The emphasis of this article is that the authors of publications are the final holders of knowledge and expertise. It is difficult to use an organization as a provider of knowledge; even the Scopus system itself warns that determining affiliation is challenging, since authors may write the names of their institutions differently. More realistic is to identify the organizations involved in topics through profiles of highly cited authors.

The two most-cited publications that have Indian co-authors are given in Table 19.

Table 19. Examples of the most cited publications by Authors from India and their co-authors, Authors ID, Title and DOI.

Authors	Author(s) ID	Title. DOI	Cited by
Kamble S.S., Gunasekaran A., Gawankar S.A.	36864045000, 56238759300, 55846962200	Sustainable Industry 4.0 framework: A systematic literature review identifying the current trends and future perspectives. [21]	464
Kamble S.S., Gunasekaran A., Sharma R.	36864045000, 56238759300, 57196545771	Analysis of the driving and dependence power of barriers to adopt industry 4.0 in Indian manufacturing industry. [22]	229

Profiles of Indian co-authors: Gawankar, Shradha Ashok; National Institute of Industrial Engineering, Mumbai, India; Scopus ID 55846962200; 10 Documents by author; 793 Citations by 762 documents; 7 h-index

Sharma, Rohit; National Institute of Industrial Engineering, Mumbai, India; Scopus ID 57196545771; https://orcid.org/0000-0001-7160-4862; 19 Documents by author; 830 Citations by 776 documents; 8 h-index

Profiles of co-authors from other countries: Gunasekaran, Angappa; California State University, Los Angeles, Los Angeles, United States; Scopus ID 56238759300; 571 Documents by author; 33055 Citations by 22776 documents; 96 h-index. Kamble, Sachin S.; EDHEC Business School, Roubaix, France; Scopus ID 36864045000 https://orcid.org/0000-0003-4922-8172; 97 Documents by author; 3305 Citations by 2688 documents; 25 h-index

The data show that non-Indian co-authors have higher citation rates, which may affect the citation of the articles in the table.

I think it would be worthwhile to conduct a separate study of **the reasons why India's research performance is high in the area of "Digitalization of Industry"**. Advancing research topics is important both for attracting funding and for gaining additional scientific knowledge through broader collaboration.

In addition to the term industry 4.0, the titles of the cited publications include Supply Chain Disruptions, so a possible theme for a more detailed study could be: **Supply Chain Disruptions**

and Industry 4.0.

Russia, within this cluster, has the greatest cooperation with India, Malaysia, Saudi Arabia, and Turkey.

Similar results, but for the second cluster, are presented in Table 20.

Table 20. Top Countries by number of publications (Occurrences) in cluster #2

Country	cluster	Occurrences	Avg. pub. year	Avg. norm. citations
italy	2	599	2020.3389	1.29
portugal	2	156	2020.4487	1.1906
poland	2	150	2020.42	0.8464
taiwan	2	145	2020.131	1.0742
austria	2	130	2020.1769	1.6275
iran	2	129	2020.3798	1.2064
pakistan	2	129	2020.6357	1.2364
indonesia	2	128	2019.9219	0.5852
hungary	2	76	2020.2763	1.0597
romania	2	76	2020.0132	0.7591

Here, publications with authors from Italy dominate. From the total table of publications, two articles satisfying the query "… where Affiliations like '%Italy%'" are presented in Table 21.

Table 21. Examples of the most cited publications by Authors from Italy and their co-authors, Authors ID, Title and DOI.

Authors	Author(s) ID	Title. DOI	Cited by
Bressanelli G., Adrodegari F., Perona M., Saccani N.	57194720467, 55932417100, 7003407947, 15849430200	Exploring how usage-focused business models enable circular economy through digital technologies. [29]	210
Chiarello F., Trivelli L., Bonaccorsi A., Fantoni G.	57215031142, 57202071072, 7005502691, 8515018000	Extracting and mapping industry 4.0 technologies using Wikipedia. [30]	101

The first publication is more concrete and has twice the citation rate, thus more suitable for identifying issues worthy of more detailed research.

Profiles of co-authors**:**

Bressanelli, Gianmarco; Università degli Studi di Brescia, Brescia, Italy; Scopus ID 57194720467 https://orcid.org/0000-0003-2255-9983; 18 Documents by author; 568 Citations by 487 documents; 8 h-index.

Adrodegari, Federico; Università degli Studi di Brescia, Brescia, Italy; Scopus ID 55932417100; https://orcid.org/0000-0001-7939-4150; 33 Documents by author; 771 Citations by 658 documents; 13 h-index.

Perona, Marco; Università degli Studi di Brescia, Brescia, Italy; Scopus ID 7003407947; https://orcid.org/0000-0002-9082-2950; 57 Documents by author; 1901 Citations by 1677 documents; 22 h-index.

Saccani, Nicola; Università degli Studi di Brescia, Brescia, Italy; Scopus ID 15849430200; https://orcid.org/0000-0002-4603-5060; 76 Documents by author; 2607 Citations by 1963 documents; 25 h-index.

The authors are from the same university, which contrasts with the variant for Indian authors discussed above.

The goal of a prospective study may be as follows: **development of the circular economy through digital technology**.

The countries in the second cluster have significantly more publications than those in the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2022 doi:10.20944/preprints202210.0044.v1

first cluster. The results are presented in Table 22.

Cluster 3

Table 22. Top Countries by number of publications (Occurrences) in cluster #3

Country	cluster	Occurrences	Avg. pub. year	Avg. norm. citations
china	3	1274	2020.4733	1.309
united kingdom	3	1167	2020.1602	1.4262
germany	3	672	2020.1771	1.3402
australia	3	631	2020.1775	1.1918
south korea	3	271	2020.0443	1.0527
sweden	3	214	2020.1963	1.2387
netherlands	3	207	2020.0918	1.2115
switzerland	3	178	2020.118	1.3937
norway	3	157	2020.5159	1.2094
finland	3	155	2020.129	1.8575

China and Great Britain have similar figures. Here are examples of highly cited papers for each country in Table 23.

Table 23. Examples of the most cited publications by Authors from China and their co-authors, Authors ID, Title and DOI.

Authors	Author(s) ID	Title. DOI	Cited by
Tao F., Cheng J., Qi Q., Zhang M., Zhang H., Sui F.	12141248300, 57192081753, 57193673728, 57192100009, 57190839107, 57189511188	Digital twin-driven product design, manufacturing and service with big data. [10]	1059 China
De Reuver M., Sørensen C., Basole R.C.	35776955900, 7102879062, 24329464200	The digital platform: A research agenda. [8]	500 United Kingdom

A high number of co-authors is typical for Chinese publications. This can increase the citation rate of a publication, since each author may have a non-overlapping set of colleagues who could be interested in such a work.

The publication [10] was previously considered in the context of clustering based on co-authorship and the topic: “**Digital twin-driven product design and manufacturing**” was treated as reasonable for a more detailed analysis, so I will focus on the English publication and its co-authors.

Authors Profiles in Scopus:

Sørensen, Carsten; London School of Economics and Political Science, London, United Kingdom; Scopus ID 7102879062; https://orcid.org/0000-0002-2002-9383; 67 Documents by author; 2833 Citations by 2243 documents; 21 h-index.

Reuver, Mark De; Faculteit Techniek, Bestuur en Management, TU Delft, Delft, Netherlands; Scopus ID 35776955900; https://orcid.org/0000-0002-6302-7185; 111 Documents by author; 2135 Citations by 1869 documents; 23 h-index.

Basole, Rahul C.; Accenture, New York, United States; Scopus ID 24329464200; 80 Documents by author; 2434 Citations by 1893 documents; 22 h-index.

There is only one author from England in this article. Northern Europe and North America are characterized by publications with international co-authors. In conjunction with the large number of publishers of scientific papers in these countries, articles by such authors, along with their quality, are well cited, meaning that their topics are highly branded, a factor important for attracting resources for the ongoing research.

Accenture is a global professional services company with leading capabilities in digital, cloud and security, and its authors' participation underlines the importance of the publication for the industry.

The topic “Digital Platform” is very broad. The term “digital platform” appears 591 times

 doi:10.20944/preprints202210.0044.v1

in the metadata I used to analyze. You can make the task more specific, for example, by adding the terms product design or manufacturing.

Data for the last fourth cluster are given in Table 24.

Cluster 4

Table 24. Top Countries by number of publications (Occurrences) in cluster #4

Country	cluster	Occurrences	Avg. pub. year	Avg. norm. citations
united states	4	1632	2019.894	1.1957
spain	4	444	2020.2297	0.8859
brazil	4	253	2020.2925	1.5445
denmark	4	140	2020.4357	1.1693
japan	4	134	2019.9179	1.2473
south africa	4	109	2020.4037	1.9153
mexico	4	64	2020.1094	0.7644
thailand	4	54	2019.9074	0.5506
viet nam	4	51	2020.6275	1.6194
nigeria	4	48	2020.5	1.4158

In this cluster, the United States leads by a considerable margin. So, I limit my attention to publications related to it, see Table 25. An interesting observation is that South Africa has the highest Average normalization citations. Some reasons for this are that only the best papers from developing countries are funded for publication in ranked journals or that a small number of authors from such countries reveal their expertise, which is of interest but rarely represented in publications by authors from developed countries.

Table 25. Examples of the most cited publications by U.S. Authors and their co-authors, Authors ID, Title and DOI.

Authors	Author(s) ID	Title. DOI	Cited by
De Reuver M., Sørensen C., Basole R.C.	35776955900, 7102879062, 24329464200	The digital platform: A research agenda. [8]	500
Kamble S.S., Gunasekaran A., Gawankar S.A.	36864045000, 56238759300, 55846962200	Sustainable Industry 4.0 framework: A systematic literature review identifying the current trends and future perspectives. [21]	464
Li L.	55760654300	China’s manufacturing locus in 2025: With a comparison of “Made-in-China 2025” and “Industry 4.0”. [31]	378

The first paper in this table was already considered in cluster 3, so I will discuss the other two.

Author profiles in Scopus of the second paper:

Gunasekaran, Angappa; California State University, Los Angeles, United States; Scopus ID 56238759300; 571 Documents by author; 33121 Citations by 22815 documents; 96 h-index.

Kamble, Sachin S.; EDHEC Business School, Roubaix, France; Scopus ID 36864045000; https://orcid.org/0000-0003-4922-8172; 98 Documents by author; 3335 Citations by 2705 documents; 25 h-index.

Gawankar, Shradha Ashok; National Institute of Industrial Engineering, Mumbai, India; Scopus ID 55846962200; 10 Documents by author; 800 Citations by 769 documents; 7 h-index.

The publication [21] is of international staff. Note that Gunasekaran, Angappa is of Indian descent. Involving compatriots in collaborative research is a significant potential for emerging economies.

Author profiles in Scopus of the last paper:

Li, Ling; Old Dominion University, Norfolk, United States; Scopus ID 55760654300; 127 Documents by author; 6111 Citations by 5009 documents; 39 h-index

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2022 doi:10.20944/preprints202210.0044.v1

In this article, the author compares the goals and achievements of China, Germany, and the United States in the digital industry, in particular.

After reading this article, the following objectives are put forward for in-depth analysis: **challenges, strategic plans, and implementations for the development of Digital Industry**.

Index Keywords

To use Index Keywords for direct choice of research topics is not rational. This issue was previously discussed on the example of the Data Storage Index Keyword. Our task, however, is different; we select authors with numerous publications and high-citation rate, whose articles are assigned to a certain category, in this case determined by the co-occurrence of the Index Keywords. Index Keywords are in fact Elsevier's controlled vocabulary, which, together with their large amount for each publication, yields a few stable clusters. Stability can be tested by changing the clustering parameters within small limits. In this case, we have four clusters.

The selection of keywords for identifying promising publications is possible according to different parameters, in this case (Table 26) the 10 most frequent Index Keywords and the 5 with the maximum Average normalization citations were used.

Table 26. Top Index Keywords by the number of publications in whose metadata they appear (Occurrences) in cluster #1.

Index Keywords	cluster	Occurrences	Avg. pub. year	Avg. norm. citations
human	1	685	2019.9781	0.9672
humans	1	449	2019.9889	1.0826
procedures	1	203	2019.4877	1.2717
priority journal	1	169	2019.1124	1.7594
adult	1	157	2019.8089	0.7131
female	1	149	2019.8322	0.6319
controlled study	1	142	2019.9366	0.5804
male	1	135	2019.8148	0.6388
internet	1	133	2019.6617	1.743
health care	1	124	2020.0806	1.1681
betacoronavirus	1	10	2020.2	6.7188
pneumonia, viral	1	11	2020.1818	6.6295
virus pneumonia	1	11	2020.1818	6.6295
coronavirus infection	1	13	2020.1538	5.6288
coronavirus infections	1	13	2020.1538	5.6288

I used SQL like 'human' operator to select papers from the full metadata table. The two most cited are in table 27.

Table 27. Examples of the most cited publications with the Index Keyword human in their metadata and their Authors, Authors ID, Title and Index Keywords list.

Authors	Authors ID	Title.	Cit	Index Keywords
Kamble S.S., Gunasekaran A., Gawankar S.A.	36864045000, 56238759300, 55846962200,	Sustainable Industry 4.0 framework: A systematic literature review identifying the current trends and future perspectives	464	Accident prevention; Augmented reality; Big data; Flow control; Internet of things; Sustainable development; Automated manufacturing environments; **Human** machine interaction; Organizational performance; Process safety; Research and development; Smart manufacturing; Sustainable industries; Systematic literature review; Industry 4.0
Li L.	55760654300,	China's	378	Embedded systems; Industrial

		manufacturing locus in 2025: With a comparison of "Made-in-China 2025" and "Industry 4.0"		research; Industry 4.0; Internet of things; Investments; Collaborative robots; Cyber-Physical System (CPS); Emerging economies; Human capitals; Internet of Things (IOT); Made in China; Manufacturing capability; Industrial economics; algorithm; automation; comparative study; **human** capital; manufacturing; research and development; robotics; China; Germany
Tseng M.-L., Tan R.R., Chiu A.S.F., Chien C.-F., Kuo T.C.	44261803600, 56962727300, 57215070441, 57219656467, 7401533845,	Circular economy meets industry 4.0: Can big data drive industrial symbiosis?	203	Big data; Digital storage; Supply chains; Circular economy; Data-driven analysis; Economic perspective; Industrial 4.0; Industrial networks; Industrial symbiosis; Multiple supplies; Optimization solution; Industrial economics; automation; economic analysis; industrial production; model test; optimization; sustainability; economic aspect; **human**; symbiosis
Özdemir V., Hekim N.	57203613602, 57555263500,	Birth of Industry 5.0: Making Sense of Big Data with Artificial Intelligence, "the Internet of Things" and Next-Generation Technology Policy	138	Article; artificial intelligence; big data; biomedicine; **human**; industry; information; Internet; medicine; nonhuman; policy; priority journal; technology; legislation and jurisprudence; procedures; trends; Artificial Intelligence; Big Data; **Humans**; Internet; Technology

The dominant theme is Industry 4.0 in its usual context: artificial intelligence, big data, industry, information, Internet, data-driven analysis, economic outlook, augmented reality, Internet of Things (IOT) combined with the keyword "human". These results are expected. Interesting is the emergence of the **Industry 5.0** theme presented in the article [32].

The term "Industry 5.0" appears 128 times in the collected metadata. It is possible to trace the emergence of the term in more detail over the past years, but this is beyond the scope of this article and can be addressed in a special publication revealing **the specific aspects of Industry 5.0 digitalization that distinguish it from Industry 4.0**.

Authors profiles in Scopus**:**

Özdemir, Vural; OMICS: A Journal of Integrative Biology, New Rochelle, United States; Scopus ID 57203613602; 186 Documents by author; 3374 Citations by 2562 documents; 29 h-index. This author publishes articles on medicine and genetics, so his opinion on Industry 5.0 in the life science context is of interest.

Hekim, Nezih; Biruni Üniversitesi, Istanbul, Turkey; Scopus ID 57555263500; 22 Documents by author; 323 Citations by 310 documents; 7 h-index. This author also publishes medical articles.

The interest in Society 5.0 and Industry 5.0 topic is exemplified by the 2022 paper [33], already cited three times in Scopus.

Although betacoronavirus is the most cited topic in this cluster according to "Avg. norm.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2022 doi:10.20944/preprints202210.0044.v1

citations" but is too broad and overrides the purpose of this article – Digital Industry.

While humanities challenges of Industry 4.0 dominated the previous cluster, the engineering issues prevent in the second cluster (Table 28).

Table 28. Top Index Keywords by the number of publications in whose metadata they appear (Occurrences) in cluster #2.

Index Keywords	cluster	Occurrences	Avg. pub. year	Avg. norm. citations
industry 4.0	2	610	2020.4967	2.4021
internet of things	2	430	2020.307	2.2515
decision making	2	391	2020.243	1.3097
artificial intelligence	2	316	2020.4589	1.5163
digital transformation	2	298	2020.6141	1.7262
manufacture	2	297	2020.2054	2.1907
construction industry	2	292	2020.411	1.2523
digital technologies	2	272	2020.4669	1.3183
digital twin	2	269	2020.5019	2.8927
information management	2	250	2020.308	1.5127
flow control	2	58	2020.2931	5.3381
enabling technologies	2	53	2020.2642	4.4097
manufacturing enterprise	2	21	2020.4286	4.1141
smart manufacturing	2	111	2020.5946	3.7056
fog	2	22	2019.8636	3.6695
predictive maintenance	2	24	2020.625	3.4382
fog computing	2	30	2020.2	3.2768

My choice of the topic "fog computing" is strictly subjective. Query TITLE-ABS-KEY ("fog computing") in Scopus matches 7,713 papers, and the specialized abstract database OnePetro yielded only 2 peer-reviewed publications as of July 30, 2022.

Therefore, it is reasonable to determine in which periodicals other than the Society of Petroleum Engineers the results of research on the implementation of fog computing in the oil and gas sector are published. For instance, by analyzing the proceedings of non-peer-reviewed conferences one can find out the actual problems of **fog computing for petroleum engineering**, and then discover the solution of related issues in the publications indexed in Scopus.

Table 29 shows three papers from these two most cited ones contain the term "fog computing" not only in the Index Keywords but also in the titles.

Table 29. Examples of the most cited publications with the Index Keyword Fog computing in their metadata and their Authors, Authors ID, Title and Index Keywords list.

Authors	Authors ID	Title	Cit	Index Keywords
Naha R.K., Garg S., Georgakopoulos D., Jayaraman P.P., Gao L., Xiang Y., Ranjan R.	56841650300, 55365934100, 7003391364, 24824284200, 36133254500, 7201978796, 57197711780,	Fog computing: Survey of trends, architectures, requirements, and research directions	221	Architecture; Big data; Computer architecture; Digital storage; Fault tolerance; Fog; Internet of things; Multitasking; Resource allocation; Surveys; Taxonomies; Allocation and scheduling; Computing architecture; Computing environments; Future research directions; Internet of thing (IOT); Internet of Things (IOT); IOT applications; Microservices; **Fog computing**
Kumari A., Tanwar S., Tyagi S., Kumar N.	57203787564, 56576145100, 55549655700,	Fog computing for Healthcare 4.0 environment:	213	Biomedical equipment; Data handling; Digital storage; Fog; Hospitals; Information

	57206866080,	Opportunities and challenges		services; Internet of things; Investments; Patient treatment; Capital investment; Context aware services; Future applications; Healthcare industry; Medical Devices; Real time data collections; Storage facilities; Technological advancement; **Fog computing**
Mellado J., Núñez F.	57221293455, 15072084200,	Design of an IoT-PLC: A containerized programmable logical controller for the industry 4.0	5	Containers; Control systems; Controllers; Digital storage; **Fog computing**; Industry 4.0; Automation technology; Feedback control loops; Industrial control systems; Laboratory experiments; Programmable logical controller; Programmable logical controller (PLC); Virtual device models; Wireless interfaces; Internet of things

The first article in the table is a review article, while the second, on the contrary, focuses on a specific industry – Healthcare 4.0. Case-specific articles better reveal potential problems related to the implementation of Fog computing than review publications, thus consider the profiles of the authors of the second publication.

Authors profiles in Scopus:

Kumari, Aparna; Nirma University, Institute of Technology, Ahmedabad, India; Scopus ID 57203787564; ttps://orcid.org/0000-0001-5991-6193; 32 Documents by author; 891 Citations by 662 documents; 14 h-index.

Tanwar, Sudeep; Nirma University, Institute of Technology, Ahmedabad, India; Scopus ID 56576145100; https://orcid.org/0000-0002-1776-4651; 304 Documents by author; 5115 Citations by 2854 documents; 41 h-index.

Tyagi, Sudhanshu; Thapar Institute of Engineering & Technology, Patiala, India; Scopus ID 55549655700; ttps://orcid.org/0000-0002-2989-3098; 94 Documents by author; 3194 Citations by 1933 documents; 31 h-index.

Kumar, N.; Thapar Institute of Engineering & Technology, Patiala, India; Scopus ID 57206866080; https://orcid.org/0000-0002-8923-2208; 791 Documents by author; 20779 Citations by 13029 documents; 77 h-index.

All the authors work in India but at different institutes. They have a lot of publications and high citations, which makes it reasonable to analyze other publications on the subject affiliated with Nirma University, Institute of Technology, and Thapar Institute of Engineering & Technology and determine if they are working on a single program or project.

In Table 30 again, we meet the case of the dominance of the Index Keyword Digital Storage, which seldom appears in the titles and abstracts of the publications in question but frequently in the Index Keywords.

Table 30. Top Index Keywords by the number of publications in whose metadata they appear (Occurrences) in cluster #3.

Index Keywords	cluster	Occurrences	Avg. pub. year	Avg. norm. citations
digital storage	3	1508	2020.057	1.0618
sustainable development	3	260	2020.6231	2.016

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 October 2022 doi:10.20944/preprints202210.0044.v1

commerce	3	177	2020.0395	1.0148
innovation	3	166	2020.2651	1.3495
china	3	149	2020.3624	1.7787
sustainability	3	148	2020.4122	1.6075
costs	3	146	2020.0685	0.9817
energy efficiency	3	144	2019.9444	1.4081
energy utilization	3	144	2020.2083	1.0587
carbon dioxide	3	129	2020.0388	0.8684
environmental economics	3	26	2020.4231	**4.9271**
sustainable development goal	3	18	2021.4444	**4.5386**
environmental policy	3	16	2020.5625	4.0994
pollution control	3	19	2019.8947	3.7117
environmental protection	3	35	2020.4857	3.1108

Although Scopus labels the largest number of published papers by Digital Storage, the most cited articles in this cluster refer to Environmental Economics and the Sustainable Development Goals. Both Index keywords Digital Storage and Environmental Economics met in the last two articles in Table 31.

Table 31. Examples of the most cited publications with the Index Keyword Digital storage in their metadata and their Authors, Authors ID, Title and Index Keywords list.

Authors	Authors ID	Title	Cit	Index Keywords
Tao F., Cheng J., Qi Q., Zhang M., Zhang H., Sui F.	12141248300, 57192081753, 57193673728, 57192100009, 57190839107, 57189511188,	Digital twin-driven product design, manufacturing and service with big data	1059	Big data; Design; **Digital storage**; Industrial research; Manufacture; Product design; Cyber and physical convergence; Digital twin; Future applications; Manufacturing enterprise; Product life cycle management; Product lifecycle data; Product-life-cycle; Service; Life cycle
Naha R.K., Garg S., Georgakopoulos D., Jayaraman P.P., Gao L., Xiang Y., Ranjan R.	56841650300, 55365934100, 7003391364, 24824284200, 36133254500, 7201978796, 57197711780,	Fog computing: Survey of trends, architectures, requirements, and research directions	221	Architecture; Big data; Computer architecture; **Digital storage**; Fault tolerance; Fog; Internet of things; Multitasking; Resource allocation; Surveys; Taxonomies; Allocation and scheduling; Computing architecture; Computing environments; Future research directions; Internet of thing (IOT); Internet of Things (IOT); IOT applications; Microservices; Fog computing
Dong F., Yu B., Hadachin T., Dai Y., Wang Y., Zhang S., Long R.	45661148900, 57193681442, 57196344975, 57158064800, 55733813600, 57196348422, 8396729800,	Drivers of **carbon emission** intensity change in China	182	**Digital storage**; Emission control; Energy efficiency; Regression analysis; Carbon emission intensities; China; Energy conservation and emission reductions; Industrial sector; Product structure; Quantile regression; Structural decomposition analysis (SDA); Technical progress; Industrial emissions; carbon

				emission; energy conservation; energy efficiency; **environmental economics**; Gross Domestic Product; regression analysis; urbanization; agriculture; Article; carbon footprint; China; coal power; economic development; energy conservation; industrialization; industry; urbanization; China
Ma M., Cai W.	57192905542, 57043939900,	What drives the **carbon mitigation** in Chinese commercial building sector? Evidence from decomposing an extended Kaya identity	99	Carbon; Climate models; Decomposition; **Digital storage**; Energy efficiency; Energy utilization; Office buildings; Quality control; Building energy consumption; Carbon mitigation; Commercial building; Decomposition analysis; Kaya identity; Energy conservation; carbon; building; carbon emission; decomposition analysis; emission control; energy efficiency; **environmental assessment**; environmental economics; analytic method; Article; building industry; carbon footprint; climate change; decomposition; energy conservation; energy resource; feasibility study; priority journal; quality control; China; Torreya nucifera

The first paper has already been mentioned in Cluster 1 about the co-authors, and the second when considering Fog computing, the term Digital storage is not found in these papers. It is purely the result of assigning the label Digital storage to these publications by the publishing system, so we will not consider them in this section.

Note. The list of Index keywords can be very large compared to the list of Author keywords. In the bibliometric data I collected from Scopus, the total number of (non-unique) Author Keywords is 48324 or 5.68 per article, and Index Keywords are 113305 or 17.9 per entry. Index Keywords are available for clustering, but the topic is better determined from the original texts, and the most concise summary of the content is the title of the article.

A promising research question could be the **Drivers of carbon mitigation** [34] in the context of the Digital Industry. As the last publication in the table is a more specific case, let's examine the authors' information in more detail.

Authors profile in Scopus:

Ma, Minda; Tsinghua University, Beijing, China; Scopus ID 57192905542; https://orcid.org/0000-0002-5533-656X; 34D ocuments by author; 1628 Citations by 945 documents; 22 h-index.

Cai, Weiguang; Chongqing University, Chongqing, China; Scopus ID 57043939900; 69 Documents by author; 2064 Citations by 1333 documents; 27 h-index.

Cai, Weiguang; Chongqing University, Chongqing, China; Scopus ID 57043939900; 72 Documents by author; 2192 Citations by 1406 documents; 27 h-index.

Energy efficiency in the building sector (and thus carbon mitigation) is especially important for large data centers, where projects to recover the heat dissipated when cooling servers are in demand.

The last cluster of Index keywords, shown in Table 32, can be attributed to manufacturing based on 3d printing optimized by using AI.

Table 32. Top Index Keywords by the number of publications in whose metadata they appear (Occurrences) in cluster #4.

Index Keywords	cluster	Occurrences	Avg. pub. year	Avg. norm. citations
3d printers	4	234	2020.2137	1.696
machine learning	4	207	2020.6087	1.3437
quality control	4	150	2020	1.4176
deep learning	4	141	2020.6383	1.5361
forecasting	4	129	2020.4806	1.2191
additives	4	121	2020.3471	1.3959
gas industry	4	121	2020.1901	0.8425
image analysis	4	115	2019.9391	0.6595
algorithm	4	101	2020.0396	1.198
petroleum industry	4	93	2019.957	0.6246
deep neural networks	4	24	2021	4.1635
food industries	4	19	2020.8421	4.0852
support vector machine	4	18	2020.0556	2.5202
measurement accuracy	4	17	2020.2941	2.3512

The 3D printer is perhaps the most in-demand technology in digital manufacturing, and main areas of applications are gas, petroleum, and food sectors, as listed in the table 32.

The fact that "deep neural networks" are a highly cited topic is a quite expected result. Support vector machine – inevitably found in machine learning topics. However, it is noteworthy that derived metrics like Avg. norm. citations should be used with caution, for example, the most cited publication for the Index keyword "deep neural networks" was published in 2019 and has only 24 citations [35].

3D printing is a very broad topic; it is important to almost every industry; Table 33 shows two highly cited articles related to the construction industry.

Table 33. Examples of the most cited publications with the Index Keyword 3d printers in their metadata and their Authors, Authors ID, Title and Index Keywords list.

Authors	Authors ID	Title	Cit	Index Keywords
Delgado Camacho D., Clayton P., O'Brien W.J., Seepersad C., Juenger M., Ferron R., Salamone S.	57196234223, 39360892000, 36485213300, 6508183775, 6701841561, 16241303300, 35574901800,	Applications of additive manufacturing in the construction industry – A forward-looking review	147	Accident prevention; Compensation (personnel); Construction industry; Copying; Disaster prevention; Manufacture; Printing; Productivity; Quality assurance; Wages; **3-D printing**; Construction workforces; Conventional constructions; Experimental application; High performance material; Infrastructure construction; Interdisciplinary research; Non-structural elements; 3D printers
García de Soto B., Agustí-Juan I.,	57219474603, 56204634800, 57217925661,	Productivity of digital fabrication in construction:	144	3D printers; Automation; Concretes; Construction industry; Cost benefit analysis; Costs;

Hunhevicz J., Joss S., Graser K., Habert G., Adey B.T.	57201885481, 57193861254, 8391064400, 6602448661,	Cost and time analysis of a robotically built wall		Economic and social effects; Productivity; Robotics; Walls (structural partitions); **3-D printing**; Construction automation; Digital fabrication; Industrialized construction; Labor productivity; Robot system; Robotic construction; Fabrication

The first article is an overview, and we look at the data on the authors of the second article [36]. There are many co-authors, so we present data on the three with the highest h-index.

Authors profile in Scopus:

Habert, Guillaume; ETH Zürich, Zurich ZH, Switzerland; Scopus ID 8391064400; https://orcid.org/0000-0003-3533-7896; 166 Documents by author; 6047 Citations by 4510 documents; 40 h-index.

Adey, Bryan Tyrone; ETH Zürich, Zurich ZH, Switzerland; Scopus ID 6602448661; 141 Documents by author; 1357 Citations by 1060 documents; 19 h-index.

García de Soto, Borja; NYU Abu Dhabi, Abu Dhabi, United Arab Emirates; Scopus ID 57219474603; https://orcid.org/0000-0002-9613-8105; 63 Documents by author; 568 Citations by 478 documents; 11 h-index.

"Productivity of digital fabrication in construction" AND "3-D printing" – such a request deserves an in-depth analysis.

Table 34 presents the data of the article with a high Avg. norm. Citations – 4.1635 and containing "deep neural networks" in the Index Keywords.

Table 34. Examples of the most cited publications with the Index Keyword deep neural networks in their metadata and their Authors, Authors ID, Title and Index Keywords list.

Authors	Authors ID	Title	Cit	Index Keywords
Moreno-García C.F., Elyan E., Jayne C.	56332688500, 35226184300, 36608214500,	New trends on digitisation of complex engineering drawings	22	Classification (of information); Complex networks; Computer vision; Deep learning; **Deep neural networks;** Feature extraction; Image analysis; Image segmentation; Neural networks; Processing; Contextualisation; Convolutional neural network; Digitisation; Engineering drawing; Recognition; Engineering education

Authors profile in Scopus:

Moreno-García, Carlos Francisco; Robert Gordon University, Aberdeen, United Kingdom; Scopus ID 56332688500; https://orcid.org/0000-0001-7218-9023; 44 Documents by author; 216 Citations by 145 documents; 9 h-index.

Elyan, Eyad; Robert Gordon University, Aberdeen, United Kingdom; Scopus ID 35226184300; https://orcid.org/0000-0002-8342-9026; 59 Documents by author; 1051 Citations by 912 documents; 16 h-index. Jayne,

Chrisina; Teesside University, Middlesbrough, United Kingdom; Scopus ID 36608214500; https://orcid.org/0000-0001-7292-2109; 68 Documents by author; 759 Citations by 686 documents; 15 h-index.

Topic: **digitisation of complex engineering drawings**. This topic is not perceived as a hyped one, but as the authors of the article note, "Digitising engineering drawings is becoming increasingly important mainly due to the legacy of drawings and documents that may provide rich source of information for industries".

Conclusions

The following are prospective research issues on the topic of Digital Industry based on the analysis of Scopus bibliometric data for 2018–2022. The final formulation of the tasks is indicative, for example, in the text of the article one can find the following formulations: "Industry 4.0 technologies for industrial performance", "Industry 4.0 and innovations", "Sustainable, Innovative, Technology-efficient Industry 4.0" – from them the following query can be composed for detailed analysis: Industry 4.0 technologies AND (performance OR innovations OR efficient). In any case, such queries are nothing more than a starting point for further research of promising issues and will be refined according to the results of queries to abstract databases. The subjects of new studies themselves can be formulated without a detailed analysis of bibliometric data, but it is the latter that suggests a sufficient number of high-cited papers, their authors, and the authoritative organizations with which the authors are affiliated.

The list of promising research issues on the topic of Digital Industry, which I have chosen for a more detailed study in later works (here AND and OR are operators):

- (Social dimension) AND (Digital Industry)
- (Digital twin-driven) AND (product design OR manufacturing OR service)
- (Energy management) AND (fog computing OR edge computing)
- (Industry 4.0 technologies) AND (performance OR innovations OR efficient)
- (Circular Economy) AND Industry 4.0
- Digital twin applications in manufacturing
- India AND (Digitalization of Industry)
- (Supply Chain Disruptions) AND Industry 4.0
- Industry 5.0 vs Industry 4.0 digitalization
- (Carbon mitigation) AND (Digital Industry)

Note: fuller tables then presented in the text of this article, some JSON files that can be used to view cooccurrence graphs using the app.vosviewer.com service and the main summary graphs provided by Scopus for the metadata used are available by direct link: https://figshare.com/articles/dataset/Identification_of_promising_research_issues_in_the_Digital_Industry_topic_by_analyzing_Scopus_bibliometric_data_for_2018_2022/20431071.

References

[1] Kosciejew M. The coronavirus pandemic, libraries and information: a thematic analysis of initial international responses to COVID-19. GKMC 2021;70:304–24. https://doi.org/10.1108/GKMC-04-2020-0041.

[2] Gharaibeh L, Eriksson KM, Lantz B, Matarneh S, Elghaish F. Toward digital construction supply chain-based Industry 4.0 solutions: scientometric-thematic analysis. SASBE 2022. https://doi.org/10.1108/SASBE-12-2021-0224.

[3] Hajirasouli A, Banihashemi S, Drogemuller R, Fazeli A, Mohandes SR. Augmented reality in design and construction: thematic analysis and conceptual frameworks. CI 2022;22:412–43. https://doi.org/10.1108/CI-01-2022-0007.

[4] Muhuri PK, Shukla AK, Abraham A. Industry 4.0: A bibliometric analysis and detailed overview. Engineering Applications of Artificial Intelligence 2019;78:218–35. https://doi.org/10.1016/j.engappai.2018.11.007.

[5] Wang H, Pan Y, Luo X. Integration of BIM and GIS in sustainable built environment: A review and bibliometric analysis. Automation in Construction 2019;103:41–52. https://doi.org/10.1016/j.autcon.2019.03.005.

[6] Lee C-H, Liu C-L, Trappey AJC, Mo JPT, Desouza KC. Understanding digital transformation in advanced manufacturing and engineering: A bibliometric analysis, topic modeling and research trend discovery. Advanced Engineering Informatics 2021;50:101428. https://doi.org/10.1016/j.aei.2021.101428.
[7] van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010;84:523–38. https://doi.org/10.1007/s11192-009-0146-3.
[8] de Reuver M, Sørensen C, Basole RC. The Digital Platform: A Research Agenda. Journal of Information Technology 2018;33:124–35. https://doi.org/10.1057/s41265-016-0033-3.
[9] Popkova EG, Sergi BS. Human capital and AI in industry 4.0. Convergence and divergence in social entrepreneurship in Russia. JIC 2020;21:565–81. https://doi.org/10.1108/JIC-09-2019-0224.
[10] Tao F, Cheng J, Qi Q, Zhang M, Zhang H, Sui F. Digital twin-driven product design, manufacturing and service with big data. Int J Adv Manuf Technol 2018;94:3563–76. https://doi.org/10.1007/s00170-017-0233-1.
[11] Li D, Wang H, Qin Y, Li Z, Wei X, Qin D. Mechanism of high amplitude low frequency fluctuations in a pump-turbine in pump mode. Renewable Energy 2018;126:668–80. https://doi.org/10.1016/j.renene.2018.03.080.
[12] Dennerley C, Huang J, Nielson R, Sefton M, Triantafilis J. Identifying soil management zones in a sugarcane field using proximal sensed electromagnetic induction and gamma-ray spectrometry data. Soil Use Manage 2018;34:219–35. https://doi.org/10.1111/sum.12410.
[13] Mao Y, You C, Zhang J, Huang K, Letaief KB. A Survey on Mobile Edge Computing: The Communication Perspective. IEEE Commun Surv Tutorials 2017;19:2322–58. https://doi.org/10.1109/COMST.2017.2745201.
[14] Tao F, Sui F, Liu A, Qi Q, Zhang M, Song B, et al. Digital twin-driven product design framework. International Journal of Production Research 2019;57:3935–53. https://doi.org/10.1080/00207543.2018.1443229.
[15] Qi Q, Tao F. Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison. IEEE Access 2018;6:3585–93. https://doi.org/10.1109/ACCESS.2018.2793265.
[16] Tao F, Qi Q, Wang L, Nee AYC. Digital Twins and Cyber–Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison. Engineering 2019;5:653–61. https://doi.org/10.1016/j.eng.2019.01.014.
[17] Dalenogare LS, Benitez GB, Ayala NF, Frank AG. The expected contribution of Industry 4.0 technologies for industrial performance. International Journal of Production Economics 2018;204:383–94. https://doi.org/10.1016/j.ijpe.2018.08.019.
[18] Frank AG, Dalenogare LS, Ayala NF. Industry 4.0 technologies: Implementation patterns in manufacturing companies. International Journal of Production Economics 2019;210:15–26. https://doi.org/10.1016/j.ijpe.2019.01.004.
[19] Frank AG, Mendes GHS, Ayala NF, Ghezzi A. Servitization and Industry 4.0 convergence in the digital transformation of product firms: A business model innovation perspective. Technological Forecasting and Social Change 2019;141:341–51. https://doi.org/10.1016/j.techfore.2019.01.014.
[20] Benitez GB, Ayala NF, Frank AG. Industry 4.0 innovation ecosystems: An evolutionary perspective on value cocreation. International Journal of Production Economics 2020;228:107735. https://doi.org/10.1016/j.ijpe.2020.107735.

[21] Kamble SS, Gunasekaran A, Gawankar SA. Sustainable Industry 4.0 framework: A systematic literature review identifying the current trends and future perspectives. Process Safety and Environmental Protection 2018;117:408–25. https://doi.org/10.1016/j.psep.2018.05.009.
[22] Kamble SS, Gunasekaran A, Sharma R. Analysis of the driving and dependence power of barriers to adopt industry 4.0 in Indian manufacturing industry. Computers in Industry 2018;101:107–19. https://doi.org/10.1016/j.compind.2018.06.004.
[23] Rosa P, Sassanelli C, Urbinati A, Chiaroni D, Terzi S. Assessing relations between Circular Economy and Industry 4.0: a systematic literature review. International Journal of Production Research 2020;58:1662–87. https://doi.org/10.1080/00207543.2019.1680896.
[24] Rocca R, Rosa P, Sassanelli C, Fumagalli L, Terzi S. Integrating Virtual Reality and Digital Twin in Circular Economy Practices: A Laboratory Application Case. Sustainability 2020;12:2286. https://doi.org/10.3390/su12062286.
[25] Cimino C, Negri E, Fumagalli L. Review of digital twin applications in manufacturing. Computers in Industry 2019;113:103130. https://doi.org/10.1016/j.compind.2019.103130.
[26] Negri E, Berardi S, Fumagalli L, Macchi M. MES-integrated digital twin frameworks. Journal of Manufacturing Systems 2020;56:58–71. https://doi.org/10.1016/j.jmsy.2020.05.007.
[27] Ali O, Ally M, Clutterbuck, Dwivedi Y. The state of play of blockchain technology in the financial services sector: A systematic literature review. International Journal of Information Management 2020;54:102199. https://doi.org/10.1016/j.ijinfomgt.2020.102199.
[28] Bag S, Pretorius JHC, Gupta S, Dwivedi YK. Role of institutional pressures and resources in the adoption of big data analytics powered artificial intelligence, sustainable manufacturing practices and circular economy capabilities. Technological Forecasting and Social Change 2021;163:120420. https://doi.org/10.1016/j.techfore.2020.120420.
[29] Bressanelli G, Adrodegari F, Perona M, Saccani N. Exploring How Usage-Focused Business Models Enable Circular Economy through Digital Technologies. Sustainability 2018;10:639. https://doi.org/10.3390/su10030639.
[30] Chiarello F, Trivelli L, Bonaccorsi A, Fantoni G. Extracting and mapping industry 4.0 technologies using wikipedia. Computers in Industry 2018;100:244–57. https://doi.org/10.1016/j.compind.2018.04.006.
[31] Li L. China's manufacturing locus in 2025: With a comparison of "Made-in-China 2025" and "Industry 4.0." Technological Forecasting and Social Change 2018;135:66–74. https://doi.org/10.1016/j.techfore.2017.05.028.
[32] Özdemir V, Hekim N. Birth of Industry 5.0: Making Sense of Big Data with Artificial Intelligence, "The Internet of Things" and Next-Generation Technology Policy. OMICS: A Journal of Integrative Biology 2018;22:65–76. https://doi.org/10.1089/omi.2017.0194.
[33] Carayannis EG, Morawska-Jancelewicz J. The Futures of Europe: Society 5.0 and Industry 5.0 as Driving Forces of Future Universities. J Knowl Econ 2022. https://doi.org/https://doi.org/10.1007/s13132-021-00854-2.
[34] Ma M, Cai W. What drives the carbon mitigation in Chinese commercial building sector? Evidence from decomposing an extended Kaya identity. Science of The Total Environment 2018;634:884–99. https://doi.org/10.1016/j.scitotenv.2018.04.043.
[35] Moreno-García CF, Elyan E, Jayne C. New trends on digitisation of complex engineering drawings. Neural Comput & Applic 2019;31:1695–712. https://doi.org/10.1007/s00521-018-3583-1.

 doi:10.20944/preprints202210.0044.v1

[36] García de Soto B, Agustí-Juan I, Hunhevicz J, Joss S, Graser K, Habert G, et al. Productivity of digital fabrication in construction: Cost and time analysis of a robotically built wall. Automation in Construction 2018;92:297–311. https://doi.org/10.1016/j.autcon.2018.04.004.

Выявление перспективных задач исследований по теме Цифровая индустрия на основе анализа библиометрических данных Scopus за 2018–2022 годы

Б.Н. Чигарев
Институт проблем нефти и газа РАН
E-mail: bchigarev@ipng.ru; ORCID: 0000-0001-9903-2800

Аннотация

Статья посвящена выявлению перспективных научных задач по теме «Цифровая индустрия» на основе анализа библиометрических данных платформы Scopus за 2018-2022 годы. Для определения перспективных задач использовались названия, дополнительно аннотации, высокоцитируемых работ авторов с высокой публикационной активностью. Профили авторов в Scopus использовались для определения их аффилиации, что важно для углубленного анализа отобранных перспективных задач Цифровой индустрии. Фильтрация метаданных для выявления высокоцитируемых работ по конкретным вопросам осуществлялась по ключевым словам, предметным областям, принадлежности к странам и спонсорским организациям. Дополнительно проведена кластеризация библиометрических метаданных на основе совместной встречаемости терминов с помощью программы VOSviewer. В рамках этих кластеров определялись статьи, представляющие интерес для формулирования перспективных исследовательских задач по теме «Цифровая промышленность», которые будут подвергнуты углубленному анализу в отдельных работах.

Ключевые слова: цифровая индустрия, перспективные задачи исследований, библиометрические метаданные, высокоцитируемые публикации, Scopus, VOSviewer.