 doi:10.20944/preprints202205.0306.v1

The study of promising research areas on the Petroleum Science topic via bibliometric analysis of the 2018-2021 publications in the Journal of Petroleum Science and Engineering

Boris Chigarev

ORCID: 0000-0001-9903-2800

Abstract: This article identifies promising research areas on the PETROLEUM SCIENCE topic via bibliometric analysis of the 2018-2021 publications in the highly cited Journal of Petroleum Science and Engineering, which is included in the Journal Citation Reports Section: ENERGY & FUELS — Q2 Quartile, ENGINEERING, PETROLEUM — Q1 Quartile. Bibliometric metadata from Web of Science were used for 866 articles in 2018, 1,142 — in 2019, 1,138 — in 2020, and 1,832 in 2021. The clustering of articles was performed using the texts of the Title, Abstract, Keywords, and Keywords Plus fields. The demo version of the Lingo3G algorithm was used. For the two major clusters, the most promising research topics were determined by comparing the titles of the 350 most cited and 350 least cited articles for each year. The hypothesis that low-cited papers often have the same subject matter as high-cited papers of previous years is discussed.

Keywords: promising areas of research, bibliometric analysis, Journal of Petroleum Science and Engineering, Web of Science, clustering.

Aim of the study: Identification of promising areas of research on energy topics through bibliometric analysis of metadata from leading international abstract databases.

Objective of this paper: The study of promising research areas on the PETROLEUM SCIENCE topic via bibliometric analysis of the 2018-2021 publications in the highly ranked the Journal of Petroleum Science and Engineering, included in the Journal Citation Reports Category: ENERGY & FUELS - Q2 Quartile, ENGINEERING, PETROLEUM - Q1 Quartile.

Introduction

Scientific, technological and innovative R&D is a key aspect of government policy aimed at promoting sustainable socio-economic development.

Progress in science, technology, and innovation (STI) is essential to implementing the 2030 Agenda and its 17 Sustainable Development Goals (SDGs). In particular the GOAL 7: Affordable and clean energy [1].

The scientific and technological landscape is changing rapidly, so research is underway to develop an approach to identifying promising scientific problems. Oil and gas research is not an exception [2].

 doi:10.20944/preprints202205.0306.v1

The assessment of research topics, depending on their stage of development, can be used for various purposes, primarily to make decisions about (financial) support for promising research.

The problem is not new, the work [3] is devoted to the development of a methodology for identifying promising areas of research based on a multi-criteria evaluation of the effectiveness of scientific work.

A number of studies have been conducted for a specific research area. The goal of [4] is to propose a quantitative methodology for identifying promising research frontiers based on the bibliographic information of scientific articles and patents. The paper shows that RFs (research fronts) can be identified by clustering the main technological documents, and for promising RFs, prospectivity indices for each RF are calculated in terms of growth, impact, marketability, and knowledge-intensity.

The evaluation of research topics depending on their stage of development can be used to make decisions about (financial) support for R&D projects. In [5] the authors identified factors influencing emerging research topics at an initial stage of their development. The results show that the differences between fields are strongly pronounced, so research is needed for specific fields of knowledge.

The article [6] considers the possibility of using bibliometric analysis to study the thematic profiles of the publications of the scientific staff of the Institute to determine the most promising areas of research. These results can be used in planning research and applying for grants.

The study [7] seeks to develop a taxonomy of promising technologies and considers the following six prospects along with their respective approaches to patent analysis: key, outlier, vacant, emerging, new, and converging technologies.

In the article [8], the identification of prospective research topics augmented with a predictive model was studied. Two methods of topic identification were utilized: word bag and LDA model. An open online catalog of publications in computer science (DBLP)[1] was used.

The purpose of the study [9] was to propose a methodology for exploring promising research directions in the energy sector. Data from patents and scientific articles were collected. A clustering of documents based on bibliographic relationships was carried out. A knowledge map was created by mapping connections between research directions. Finally, promising research directions were identified.

An algorithm for determining innovative research directions in the context of digitalization was developed and tested as applied to the medical sector [10]. A predictive model based on the hype cycle was developed. This model enables to define a list of potential areas of hardware development in the medical sector. It also allows to represent this list as a set of lexemes. Based on this, the authors created an associative map with the most promising areas of research in medicine.

---

[1] The *dblp computer science bibliography* provides open bibliographic information on major computer science journals and proceedings. https://dblp.org/

 doi:10.20944/preprints202205.0306.v1

The relationship of policy decisions and implementation, as reflected in scholarly publications, is discussed in [11] which argues that the bibliometric analysis of publications on the topic of energy efficient appliances shows that policies are usually implemented before major publications on the topic, with time lags ranging from 3 to 30 years. Nevertheless, this trend is changing with the appearance of publications about new devices that appear shortly after or parallel to the creation of the policy.

This brief overview illustrates the importance and potential of analysing scientific publications in order to identify promising research. In this context, the choice of article sources for analysis and the definition of criteria for evaluating promising research are crucial issues.

Materials and methods

Bibliometric metadata from Web of Science have been used for 866 articles in 2018, 1,142 in 2019, 1,138 in 2020, and 1,832 in 2021 published in the high-cited journal Petroleum Science and Engineering, which is included in Journal Citation Reports in the areas of study: ENERGY & FUELS - Q2 Quartile, ENGINEERING, PETROLEUM - Q1 Quartile. These data are current as of May 9, 2022.

I performed the clustering of articles using the texts of the Title, Abstract, Keywords, and Keywords Plus fields. A demo version of the Lingo3G document clustering algorithm was used, available at: https://search.carrotsearch.com/#/search/web. This algorithm generates fewer clusters compared to the Lingo algorithm and better label assignment when using the default parameters.
For the two major clusters, the most promising research topics were determined by comparing the titles of the 350 most cited and 350 least cited articles for each year.

Web of Science tags used in the tables and attachments:

- AF, TI, DE, ID, AB, TC, Z9, U1
- AF Author Full Name
- TI Document Title
- DE Author Keywords
- ID Keywords Plus
- AB Abstract
- TC Times Cites in Web of Science Core Collection
- Z9 Total Times Cited Count (Web of Science Core Collection, Arabic Citation Index, BIOSIS Citation Index, Chinese Science Citation Database, Data Citation Index, Russian Science Citation Index, SciELO Citation Index)
- U1 Usage Count (Last 180 Days)

I used the following attaching file names: file-result- year-350-(low or top).csv, for example: file-result-2019-350-top.csv

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

The tables in this article are compiled from data taken from these files. In text, the Cluster Level 1 field is referred to simply as Cluster, and the Cluster Level 2 field as topic.

The results of article clustering for 2018

Figure 1 shows the results of clustering the 350 most cited articles published in 2018 by the Journal of Petroleum Science and Engineering.

To cluster the documents, I used the fields: title, abstract, author keywords, and keywords plus.

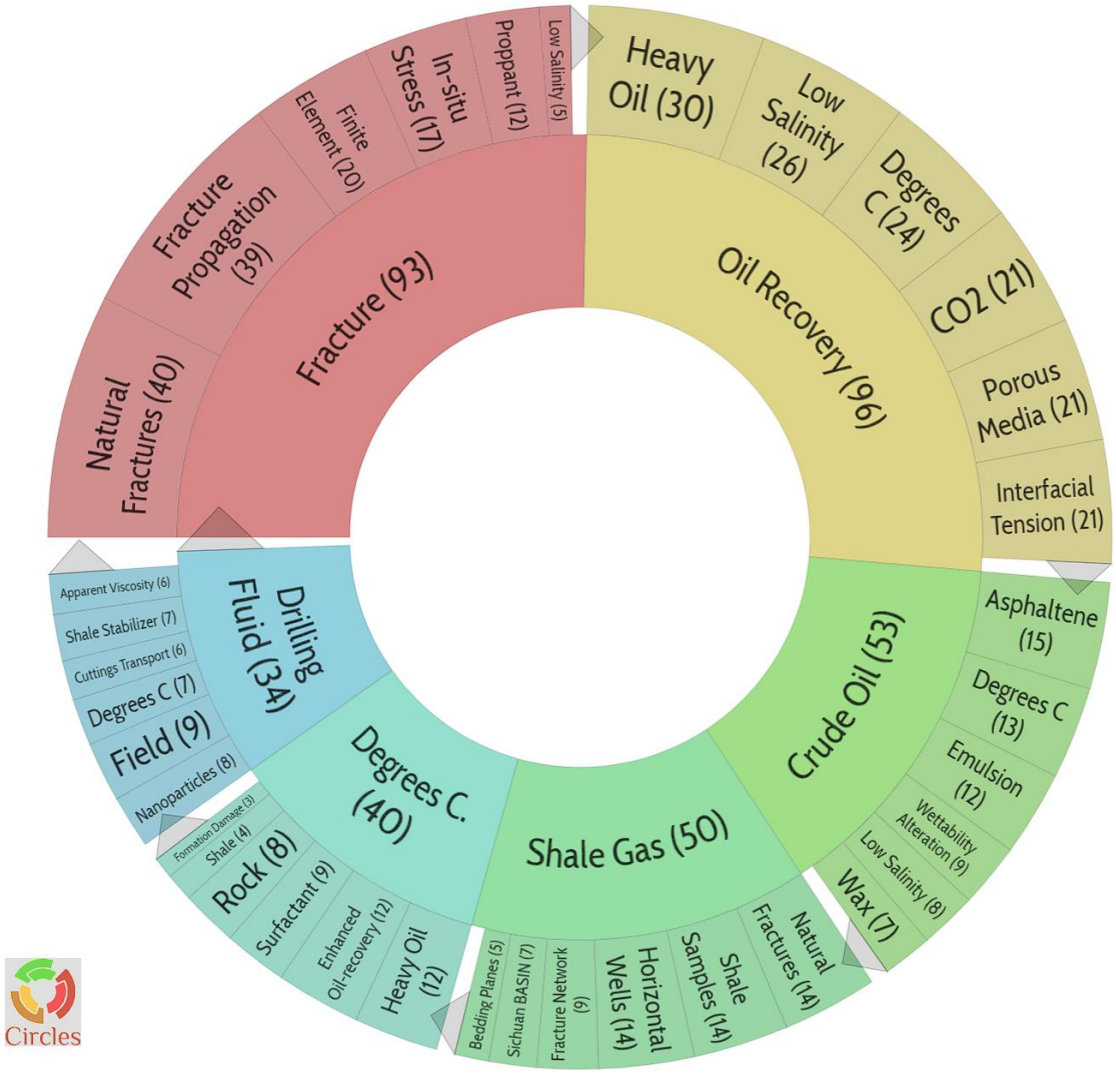

Fig. 1. Pie-chart for 350 top cited articles in 2018. A total of 28 clusters were identified, in which 91.7% of the documents fell.

Fracture (93 articles) and Oil Recovery (96 articles) are the top two first-level clusters by the number of publications.

Note: I will use the term cluster to refer to first-level clusters and the term topic to refer to second-level clusters within the cluster in question.

Heavy Oil and Low Salinity topics represent the majority of papers in the Oil Recovery cluster.

 doi:10.20944/preprints202205.0306.v1

Natural Fractures and Fracture Propagation topics represent the majority of papers in the Fracture cluster.

To provide a more comprehensive disclosure of the above topics, Tables 1 - 4 present the titles of the most cited articles.
If you have the title of the article, the year of publication, and the journal name, it is easy to find the full bibliography of the article.

Table 1. The most cited articles in the cluster level 1 Fracture and the cluster level 2 Natural Fractures

TI	TC
Numerical simulation of hydraulic fracture propagation in naturally fractured formations using the cohesive zone model	77
Analysis of hydraulic fracture initiation and propagation in deep shale formation with high horizontal stress difference	58
Wellbore stability analysis during drilling through marine gas hydrate-bearing sediments in Shenhu area: A case study	49
Role of molecular diffusion in heterogeneous, naturally fractured shale reservoirs during CO2 huff-n-puff	41

The most telling terms on this topic are numerical simulation, cohesive zone model, hydraulic fracture initiation, deep shale formation, molecular diffusion, and CO2 huff-n-puff.

It is worth noting article 3 in Table 1, which discusses drilling risks associated with natural gas hydrate-bearing sediments, which can lead to wellbore instability and fractures in the formation.

Table 2. The most cited articles in the cluster level 1 Fracture and the cluster level 2 Fracture Propagation.

TI	TC
Influence of gravel on the propagation pattern of hydraulic fracture in the glutenite reservoir	96
Numerical simulation of hydraulic fracture propagation in naturally fractured formations using the cohesive zone model	77
Surface characteristics and permeability enhancement of shale fractures due to water and supercritical carbon dioxide fracturing	64
Experimental investigation of quenching effect on mechanical, microstructural and flow characteristics of reservoir rocks: Thermal stimulation method for geothermal energy extraction	60

The most telling terms in Table 2 are propagation pattern of hydraulic fracture, numerical simulation of hydraulic fracture propagation, naturally fractured formations, permeability enhancement of shale fractures, mechanical, microstructural and flow characteristics of reservoir rocks.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

Note that the article "Numerical simulation of hydraulic fracture propagation in naturally fractured formations using the cohesive zone model" appears in both topics of the Fracture cluster.

Heavy Oil and Low Salinity topics related to the Oil Recovery cluster are represented by the titles of the most cited articles in Tables 3 and 4.

Table 3. The most cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Heavy Oil

TI	TC
Flow simulation of the mixture system of supercritical CO2 & superheated steam in toe-point injection horizontal wellbores	59
Application of high resolution NMR (H-1 and C-13) and FTIR spectroscopy for characterization of light and heavy crude oils	38
Stability and flooding analysis of nanosilica/ NaCl /HPAM/SDS solution for enhanced heavy oil recovery	37
Experimental study of nanoparticle and surfactant stabilized emulsion flooding to enhance heavy oil recovery	36

Flow simulation, supercritical CO2 & superheated steam, nanosilica/ NaCl /HPAM/SDS, nanoparticle and surfactant stabilized emulsion, high resolution NMR (H-1 and C-13), and FTIR spectroscopy - these terms are the most typical for highly cited papers on enhanced heavy oil recovery.

Table 4. The most cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Low Salinity

TI	TC
An experimental and modeling study to investigate brine-rock interactions during low salinity water flooding in carbonates	55
Low salinity waterflooding for a carbonate reservoir: Experimental evaluation and numerical interpretation	47
Carbon dioxide/water foams stabilized with a zwitterionic surfactant at temperatures up to 150 degrees C in high salinity brine	46
pH effect on wettability of oil/brine/carbonate system: Implications for low salinity water flooding	43
Direct pore-scale visualization of interactions between different crude oils and low salinity brine	40

Brine-rock interactions, low salinity water flooding in carbonates, pH effect on wettability, and direct pore-scale visualization are the main issues addressed in the highly cited articles on low-salinity water flooding.

Figure 2 shows the clustering of the 350 least-cited articles published in the Journal of Petroleum Science and Engineering in 2018.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

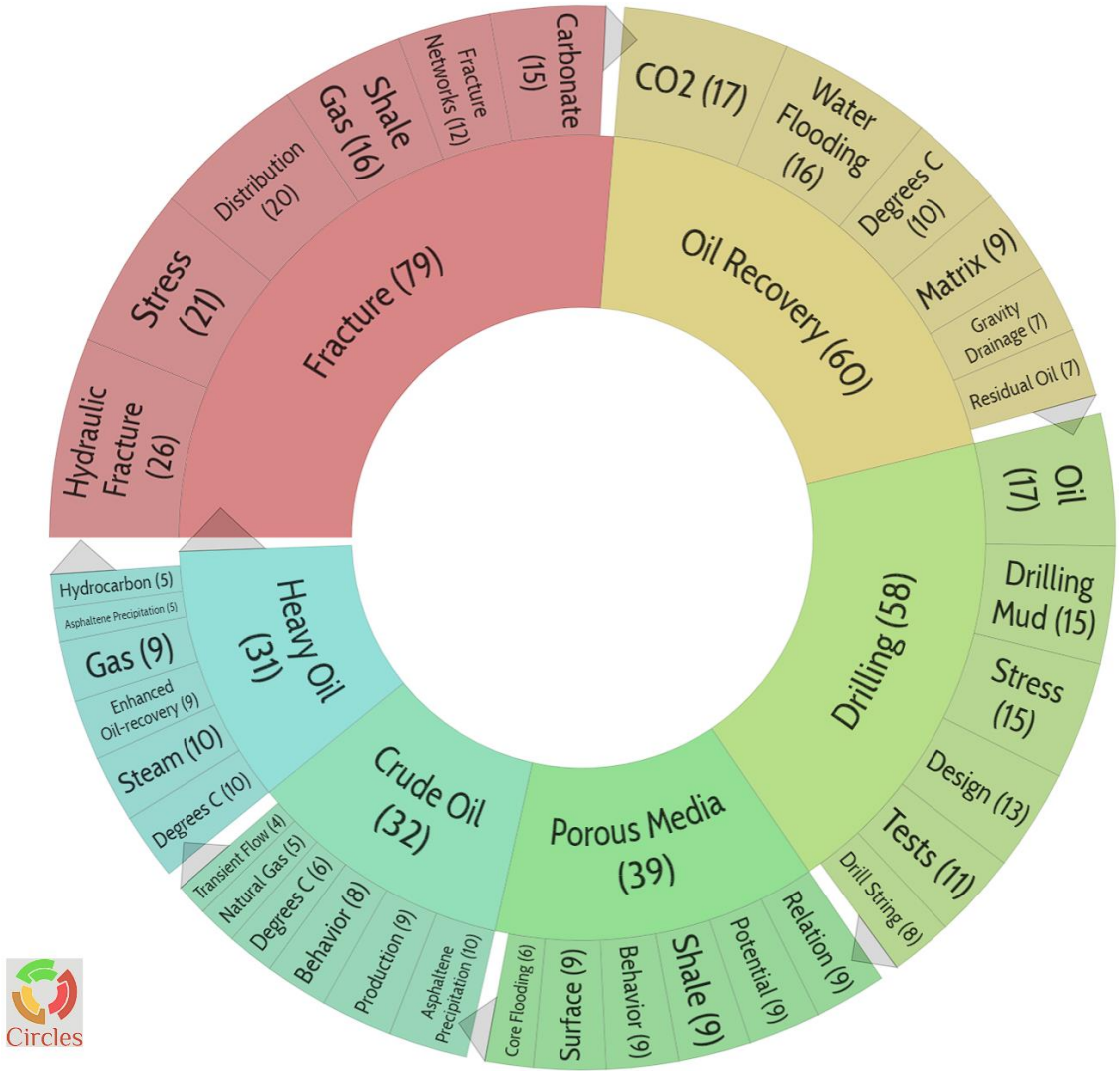

Fig. 2. Pie-chart for 350 least-cited articles in 2018. A total of 28 clusters were identified, in which 84.5% of the documents fell.

As in Figure 1, the two main clusters are Fracture (79) and Oil Recovery (60), but the leading themes within these clusters are different. In the first cluster it is Hydraulic Fracture and Stress, and in the second cluster it is CO2 and Water Flooding.

The least-cited publications for these two clusters and their topics are presented in Tables 5-8.

Table 5. The least-cited articles in the cluster level 1 Fracture and the cluster level 2 Hydraulic Fracture

TI	TC
Empirical modeling of two-phase CBM production using analogy to nature	2
Managed Saffman-Taylor instability during overflush in hydraulic fracturing	2
A generalization to transient bilinear flows	2
Mineralogical variability of the Permian Roseneath and Murteree Shales from the Cooper Basin, Australia: Implications for shale properties and hydrocarbon extraction	2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

The terms empirical modeling, Saffman-Taylor instability, overflush in hydraulic fracturing, transient bilinear flows, mineralogy variability and shale properties describe the issues associated with hydraulic fracturing.

Table 5 shows the older tasks for Hydraulic Fracture compared to Table 1, which focuses on Natural Fractures.

Table 6. The least-cited articles in the cluster level 1 Fracture and the cluster level 2 Stress

TI	TC
Modeling and analysis of coupled thermal-hydrologic-mechanical processes during lost circulation	0
Localized stress field modelling around fractures using three-dimensional discrete element method	1
Managed Saffman-Taylor instability during overflush in hydraulic fracturing	2
Mineralogical variability of the Permian Roseneath and Murteree Shales from the Cooper Basin, Australia: Implications for shale properties and hydrocarbon extraction	2

Coupled thermal-hydrologic-mechanical processes, localized stress field modelling, and three-dimensional discrete element method are terms that describe the tasks of determining stress states.

The last two items in both tables are the same, indicating a high correlation between the topics Hydraulic Fracture and Stress. As opposed to Table 2, Table 6 focuses on determining the stress state rather than crack propagation.

Table 7. The least-cited articles in the cluster level 1 Oil Recovery and the cluster level 2 CO2

TI	TC
A new hybrid production optimization algorithm for the combined CO2 -cyclic solvent injection (CO2 -CSI) and water/gas flooding in the post-CHOPS reservoirs	1
Comparative study of oil spreading characteristics for water and carbonated water systems using live and dead oils	2
Possibility of decreasing CO2 emissions from flaring on a mature oil field	3
Analytical approach for leakage characterization in carbon sequestration in a bounded deep saline aquifer	4

Combined CO2-cyclic solvent injection, carbonated water systems, CO2 emissions, and leakage characterization in carbon sequestration are common tasks in CO2 Oil Recovery.

Table 8. The least-cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Water Flooding

TI	TC
A new experimental method for measuring the three-phase relative permeability of oil, gas, and water	1

 doi:10.20944/preprints202205.0306.v1

TI	TC
A new hybrid production optimization algorithm for the combined CO2-cyclic solvent injection (CO2-CSI) and water/gas flooding in the post-CHOPS reservoirs	1
Elasticity and electrical resistivity of chalk and greensand during water flooding with selective ions	3
An experimental study to determine suitable injection strategies for water-alternating-solvent process in green and brownfields	3

experimental method, three-phase relative permeability, combined CO2-cyclic solvent injection, water flooding with selective ions, water-alternating-solvent process - these terms describe well the issues that arise in Oil Recovery by Water Flooding. The paper on combined CO2-cyclic solvent injection can serve as the link between the topics in Tables 7 and 8.

Comparing the most and least cited publications for 2018 shows that their topics are broadly similar, except that the more cited publications focus on newer issues related to Oil Recovery through Fracture.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

The results of article clustering for 2019

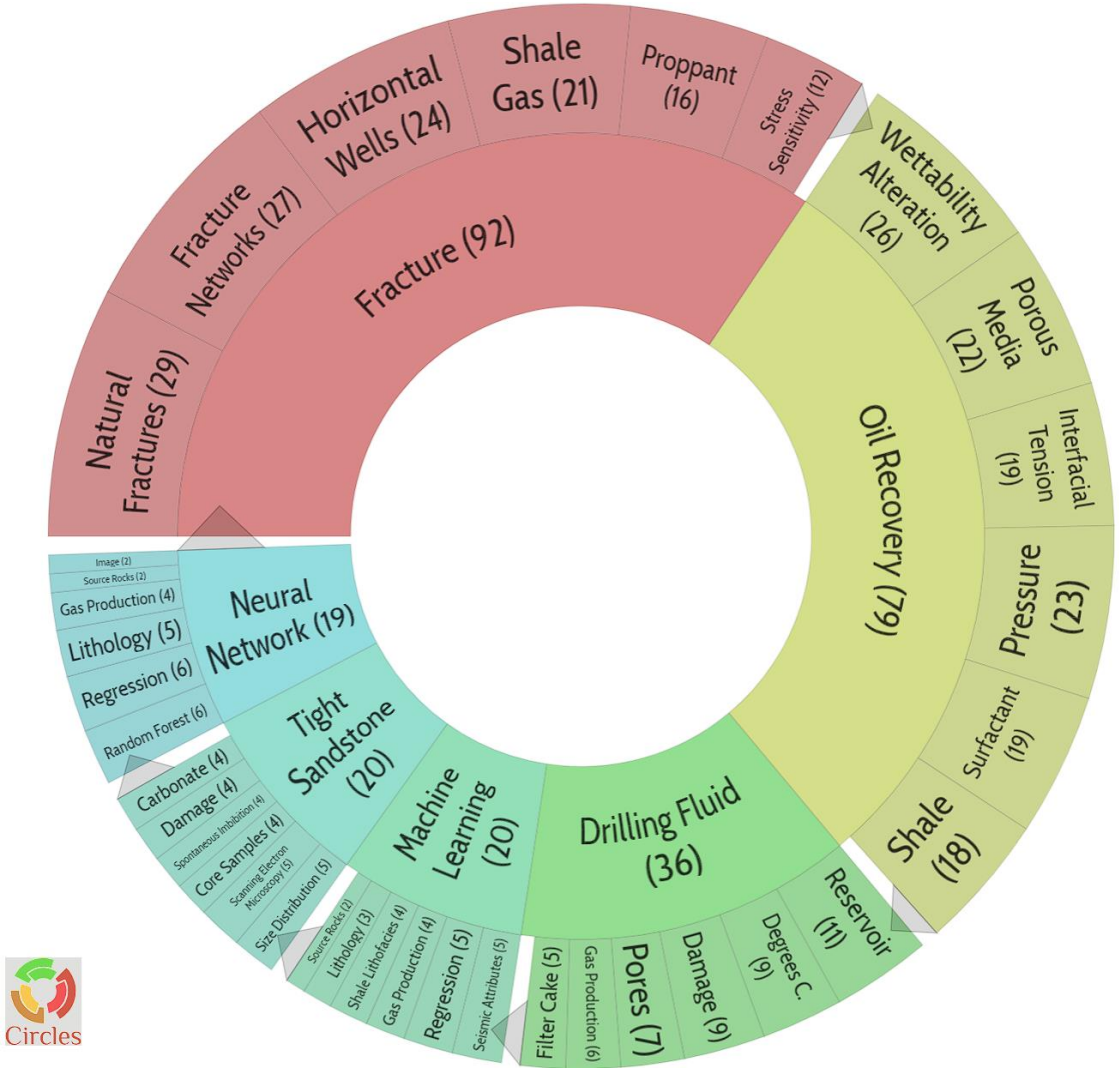

Fig. 3. Pie-chart for 350 top cited articles in 2019. A total of 28 clusters were identified, in which 84.0% of the documents fell.

Fracture and Oil Recovery dominate even more than in 2018. Of interest: Neural Network, Tight Sandstone, and Machine Learning clusters appeared. In addition to Natural Fractures, a Fracture Networks topic emerged. The Oil Recovery cluster became dominated by Wettability Alteration and Porous Media topics.

The topics of Neural Network and Machine Learning require a particular publication, especially in terms of their application area. It will be particularly interesting to discover where these methods are applied in Lithology.

Analogously to 2018, Tables 9 - 12 present the titles of the most and least cited publications.

Table 9. The most cited articles in the cluster level 1 Fracture and the cluster level 2 Natural Fractures

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

TI	TC
A new fractal approach for describing induced-fracture porosity/permeability/ compressibility in stimulated unconventional reservoirs	92
Wellbore stability in naturally fractured formations featuring dual-porosity/single-permeability and finite radial fluid discharge	48
A hybrid embedded discrete fracture model for simulating tight porous media with complex fracture systems	29
Quantitative investigation of fracture interaction by evaluating fracture curvature during temporarily plugging staged fracturing	28

The Fractal approach, induced-fracture porosity, simulating tight porous media, and complex fracture systems describe well the topic of Quantitative investigation of fracture.

Table 10. The most cited articles in the cluster level 1 Fracture and the cluster level 2 Fracture Networks

TI	TC
Brittleness evaluation of coal based on statistical damage and energy evolution theory	70
Wellbore stability in naturally fractured formations featuring dual-porosity/single-permeability and finite radial fluid discharge	48
Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling	45
Applicability of deep neural networks on production forecasting in Bakken shale reservoirs	35

Aside from fractured reservoirs, the most cited topics include unconventional reservoirs, shale reservoirs, and methods of machine learning analysis and deep neural networks.

Of particular note is the issue of coal brittleness - developing countries have significant coal reserves, and the topic of coal bed methane is important to them.

Table 11. The most cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Wettability Alteration

TI	TC
Adsorption analysis of natural anionic surfactant for enhanced oil recovery: The role of mineralogy, salinity, alkalinity and nanoparticles	75
Experimental study of wettability alteration and spontaneous imbibition in Chinese shale oil reservoirs using anionic and nonionic surfactants	52
Synergistic effects of nanoparticles and surfactants on n-decane-water interfacial tension and bulk foam stability at high temperature	44
Bio-based surfactant for enhanced oil recovery: Interfacial properties, emulsification and rock-fluid interactions	39

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

Adsorption analysis, natural anionic surfactant, spontaneous imbibition, wettability alteration, foam stability, nonionic surfactants, and bio-based surfactant describe well the topic of Wettability Alteration in the Oil Recovery cluster.

Table 12. The most cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Porous Media

TI	TC
Bio-based surfactant for enhanced oil recovery: Interfacial properties, emulsification and rock-fluid interactions	39
Oil recovery by spontaneous imbibition from partially water-covered matrix blocks with different boundary conditions	32
Effects of concentration and size of TiO2 nano-particles on the performance of smart water in wettability alteration and oil production under spontaneous imbibition	30
A hybrid embedded discrete fracture model for simulating tight porous media with complex fracture systems	29

Bio-based surfactant for enhanced oil recovery - is relevant to both the Wettability Alteration and the Porous Media.

Spontaneous imbibition, water-covered matrix blocks, nano-particles, smart water, fracture model, tight porous media- key issues of highly cited articles on the topic Porous Media in the Oil Recovery cluster.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

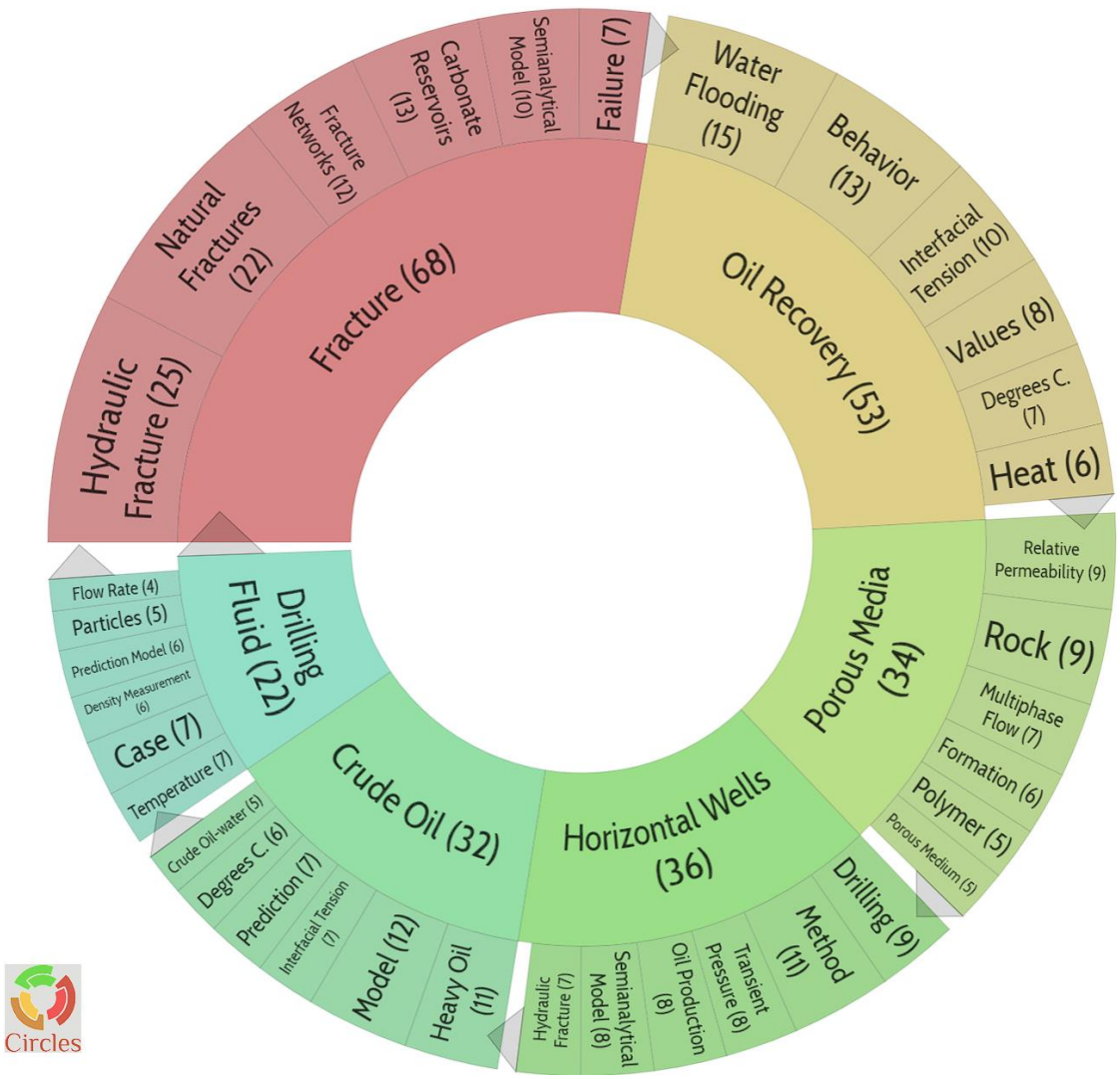

Fig. 4. Pie-chart for 350 least-cited articles in 2019. A total of 28 clusters were identified, in which 79.4% of the documents fell.

The topic Natural Fractures had a lower citation rate compared to 2018. It is consistent with the hypothesis that less-cited topics are sometimes well-cited topics from the previous period.

Hydraulic Fracture and Water Flooding are by no means new topics, perhaps that is why they are so well represented in the low-cited papers.

The highly cited topics Neural Networks and Machine Learning (Figure 3) are not represented in Figure 4.

Table 13. The least-cited articles in the cluster level 1 Fracture and the cluster level 2 Hydraulic Fracture

TI	TC
Probing hydraulically-fractured wells in unconventional shale reservoirs under cyclic CO2 injection: Variation of thermophysical properties	0

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

TI	TC
Propagation behaviors of hydraulic fractures arising from the sidewall of wellbore using high-order Generalized Finite Element Method	0
A semi-analytical model for simulation of fluid flow in tight rock with irregular fracture geometry	1
A semi-analytical model for characterizing transient flow behavior of reoriented refractures	2

Probing hydraulically-fractured, unconventional shale reservoirs, cyclic CO2 injection, Finite Element Method, fluid flow in tight rock, semi-analytical model - these are established rather than actual problems.

Table 14. The least-cited articles in the cluster level 1 Fracture and the cluster level 2 Natural Fractures

TI	TC
Azimuthally variation of elastic impedances for fracture estimation	0
Semi-analytical prediction of critical oil rate in naturally fractured reservoirs with water coning	1
Bipropellant high energy stimulation for oil and gas applications	2
Multiple failure state triaxial testing of the Montney Formation	2

The publication titles listed in Table 14 reflect tasks that have become engineering rather than actual research ones.

Table 15. The least-cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Water Flooding

TI	TC
A new approach for measuring rheology of polymer solutions in reservoir conditions	0
Experimental study of the surfactants effects on gas phase mass transfer coefficients in gas condensate reservoirs	0
Recovery efficiency of a 28 degrees API crude-oil system as a function of voidage replacement ratio	1
A mathematical model for microbial enhanced oil recovery considering the double-bacterial competition mechanism	1

Similarly, as in Table 14, measuring the rheology of polymer solutions in reservoir conditions is primarily an improvement over routine work than a new hot research topic.

Table 16. The least-cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Behavior

TI	TC
Semi-analytical transient pressure solution for power-law polymer injection in a vertical well	0

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

TI	TC
Experimental study of the surfactants effects on gas phase mass transfer coefficients in gas condensate reservoirs	0
Two-phase and three-phase equilibrium K-values for modelling of non-condensable gas Co-Injection processes	2
Analytical derivation of a pH dependent contact angle equation and the relationship to industrial processes of energy concerns	3

Reading the titles in Table 16, one would like to say that the experimental study of the effect of surfactants on gas phase mass transfer coefficients is a good topic for ongoing research, but unlikely to receive a high citation.

The results of article clustering for 2020

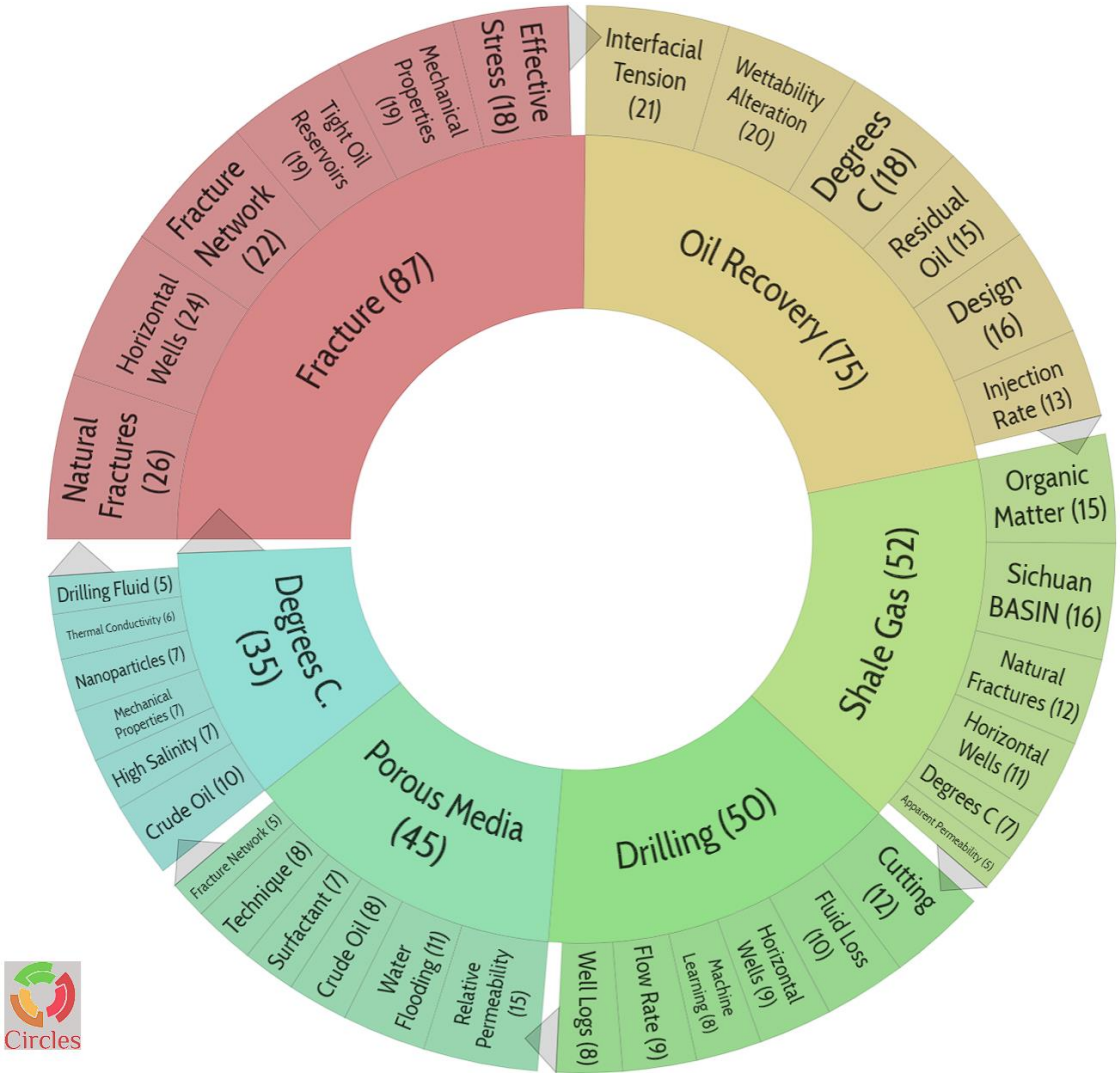

Fig. 5. Pie-chart for 350 top cited articles in 2020. Full number of clusters — 28; 90.5% clustered documents.

The titles of the articles in Tables 15-18 confirm that the topics in the 2020 publications continue to be the same as in previous years, thus making it useless to compare them in detail.

 doi:10.20944/preprints202205.0306.v1

Remark: perhaps this is related to the Covid-19 epidemic. It is just a hypothesis that requires individual consideration. Due to the possible strong influence of this factor, the data for 2020 are left without further comment - it is incorrect to analyze data without taking into account the potential influence of the factor, which may turn out to be dominant. In 2020, a great deal of attention was paid to Covid-19-related publications, and the focus of interest and, accordingly, citations were shifted to this topic. That agrees with Google Trends, which shows that there were 30 Covid-19 queries per week in 2020 and 19.5 in 2021. (https://unstats.un.org/unsd/ccsa/documents/covid19-report-ccsa_vol3.pdf)

Noticeably, the topics Neural Networks and Machine Learning left the list of most cited papers in 2020. However, the topics themselves are present in the full list of topics given in the appendix.

Table 15. The most cited articles in the cluster level 1 Fracture and the cluster level 2 Natural Fractures

TI	TC
Geological conditions and exploration potential of shale gas reservoir in Wufeng and Longmaxi Formation of southeastern Sichuan Basin, China	39
Experimental investigation into hydraulic fracture geometry and proppant migration characteristics for southeastern Sichuan deep shale reservoirs	33
Morphological and petro physical estimation of Eocene tight carbonate formation cracking by cryogenic liquid nitrogen; a case study of Lower Indus basin, Pakistan	32
Natural fractures in soft coal seams and their effect on hydraulic fracture propagation: A field study	20

Table 16. The most cited articles in the cluster level 1 Fracture and the cluster level 2 Natural Fractures

TI	TC
Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model	70
Estimation of fracture production and water breakthrough locations of multi-stage fractured horizontal wells combining pressure-transient analysis and electrical resistance tomography	29
A numerical model for fractured horizontal well and production characteristics: Comprehensive consideration of the fracturing fluid injection and flowback	20
Comprehensive experimental study of proppant transport in an inclined fracture	17

Table 17. The most cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Interfacial Tension

TI	TC
Water-oil interfacial tension (IFT) reduction and wettability alteration in surfactant flooding process using extracted saponin from Anabasis Setifera plant	37

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

TI	TC
The role of salinity and aging time on carbonate reservoir in low salinity seawater and smart seawater flooding	29
Carbonated polymeric nanofluids for enhanced oil recovery from sandstone reservoir	21
Pore-scale experimental study on EOR mechanisms of combining thermal and chemical flooding in heavy oil reservoirs	19

Table 18. The most cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Wettability Alteration

TI	TC
Water-oil interfacial tension (IFT) reduction and wettability alteration in surfactant flooding process using extracted saponin from Anabasis Setifera plant	37
The role of salinity and aging time on carbonate reservoir in low salinity seawater and smart seawater flooding	29
Experimental and simulation study of low salinity brine interactions with carbonate rocks	22
Carbonated polymeric nanofluids for enhanced oil recovery from sandstone reservoir	21

In Tables 17 and 18, 3 of the 4 articles refer to both topics.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

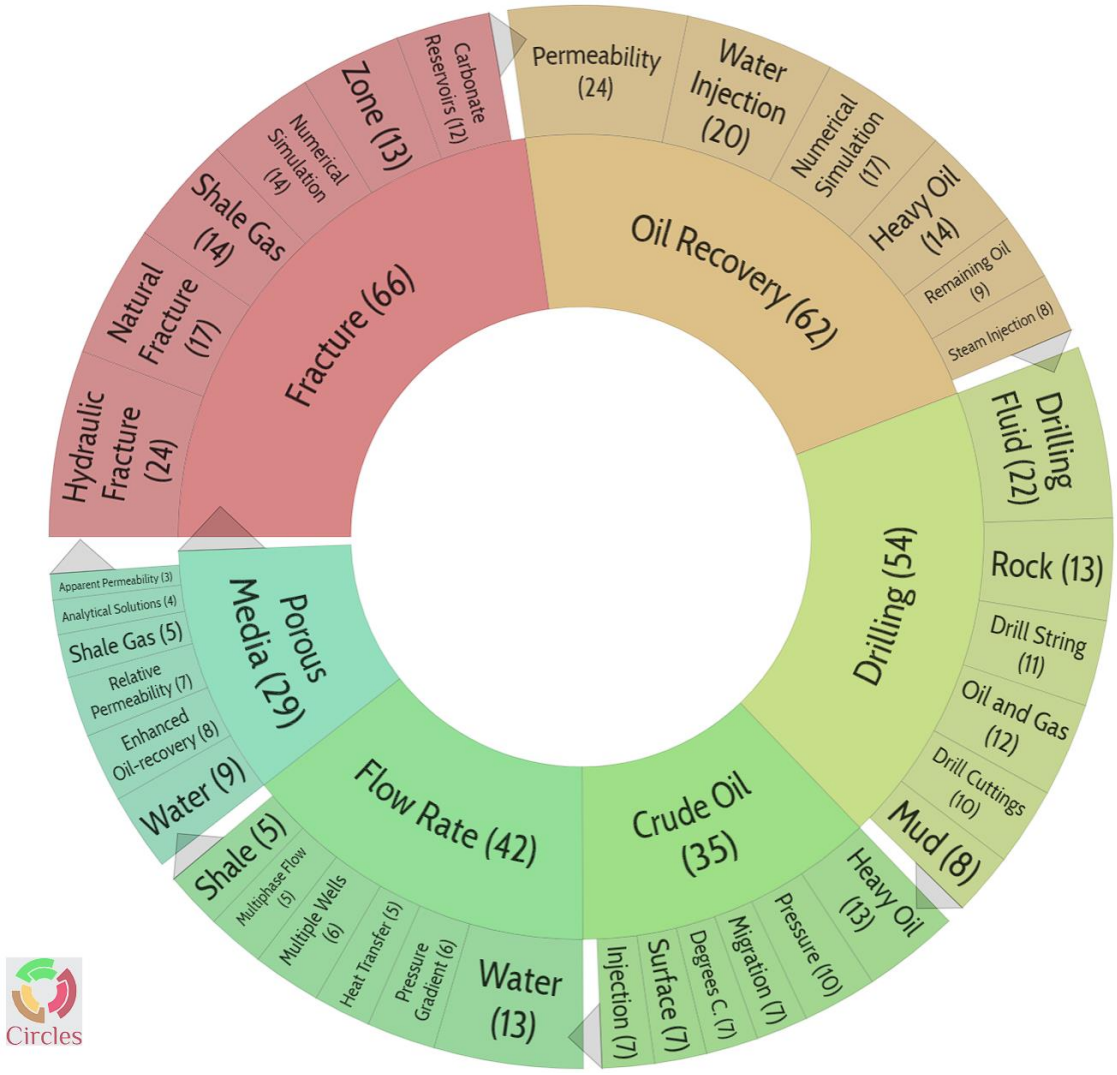

Fig. 6. Pie-chart for 350 least-cited articles in 2020. Full number of clusters — 29; 85.4% clustered documents.

The low-cited publications of 2020, like the highly cited ones, simply continue the themes of previous years, as evidenced by the titles of the articles in Tables 19-22.

Table 19. The least-cited articles in the cluster level 1 Fracture and the cluster level 2 Hydraulic Fracture

TI	TC
A semi-analytical model for simulation of multiple vertical wells with well interference	1
Pressure transient analysis in the child well with complex fracture geometries and fracture hits by a semi-analytical model	1
Fracture diagnostic using distributed temperature measurements during a pause in flow-back period	1
The use of flowback data for estimating dynamic fracture volume and its correlation with completion-design parameters: Eagle Ford cases	1

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

Table 20. The least-cited articles in the cluster level 1 Fracture and the cluster level 2 Natural Fracture

TI	TC
Mineral filling mechanism in complex carbonate reservoir fracture system : Enlightenment from numerical simulation of water-rock interaction	1
Experimental investigation and theoretical modeling of stress-dependent permeability in naturally fractured tight gas reservoir	1
Hydraulic fracturing induced fault slip and casing shear in Sichuan Basin: A multi-scale numerical investigation	2
A pore-scale experimental study of non-aqueous foam for improving hydrocarbon miscible flooding	2

Table 21. The least-cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Permeability

TI	TC
Performance assessment and forecasting of cyclic gas injection into a hydraulically fractured well using data analytics and machine learning	1
Mechanism study of the relation between the performance of CO2 immiscible flooding and rock permeability	2
Recovery efficiency of tight oil reservoirs with different injection fluids: An experimental investigation of oil-water distribution feature	2
Experimental study of small-sized polymeric microgel (SPM) in low- or median-permeability reservoirs	2

Table 22. The least-cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Permeability

TI	TC
Effects of methanol and acetone as mutual solvents on wettability alteration of carbonate reservoir rock and imbibition of carbonated seawater	1
Role of viscous cross-flow and emulsification in recovery of bypassed oil during foam injection in a microfluidic matrix-fracture system	1
Recovery efficiency of tight oil reservoirs with different injection fluids: An experimental investigation of oil-water distribution feature	2
Experimental study of small-sized polymeric microgel (SPM) in low- or median-permeability reservoirs	2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

The results of article clustering for 2021

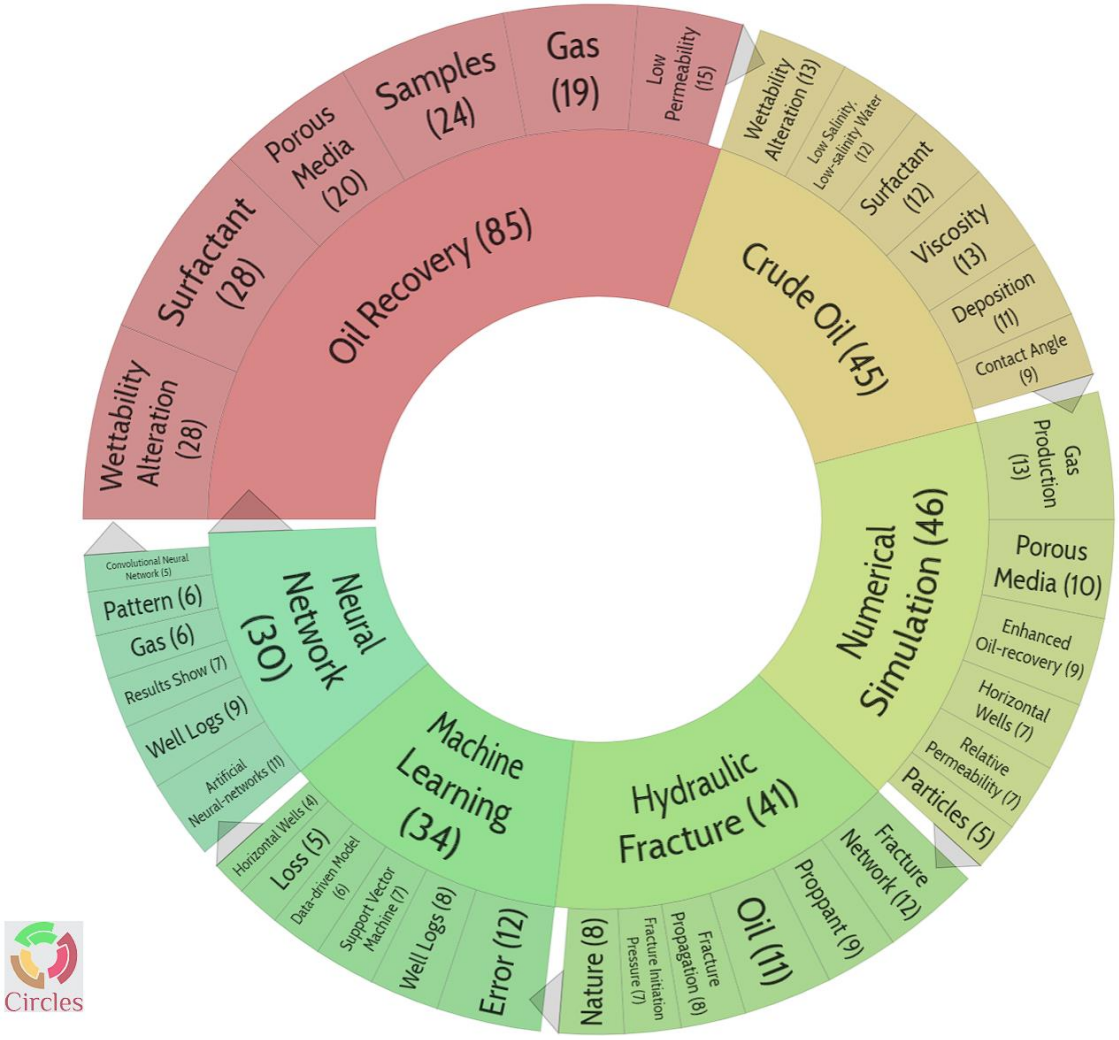

Fig. 7. Pie-chart for 350 top cited articles in 2021. Full number of clusters — 29; 86.8% clustered documents

Oil Recovery in 2021 is still the dominant cluster. Wettability Alteration and Surfactant are connected topics. Of interest in the Hydraulic Fracture cluster is the Proppant topic.

As compared to pre-Covid 2019, clusters of Neural Networks, Machine Learning, and Numerical Simulation hold a more prominent position among highly cited papers.

Table 23. The most cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Wettability Alteration

TI	TC
Simultaneous evaluation of capillary pressure and wettability alteration based on the USBM and imbibition tests on carbonate minerals	18
An introductory investigation of a polymeric surfactant from a new natural source in chemical enhanced oil recovery (CEOR)	15

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

TI	TC
The effect of brine salinity and oil components on dynamic IFT behavior of oil-brine during low salinity water flooding: Diffusion coefficient, EDL establishment time, and IFT reduction rate	11
Investigation into fluid-fluid interaction phenomena during low salinity waterflooding using a reservoir-on-a-chip microfluidic model	9

Table 24. The most cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Surfactant

TI	TC
Rheological analysis and EOR potential of surfactant treated single-step silica nanofluid at high temperature and salinity	18
Simultaneous evaluation of capillary pressure and wettability alteration based on the USBM and imbibition tests on carbonate minerals	18
An introductory investigation of a polymeric surfactant from a new natural source in chemical enhanced oil recovery (CEOR)	15
Nanoparticles as foam stabilizer: Mechanism, control parameters and application in foam flooding for enhanced oil recovery	8

The relationship between the Wettability Alteration and Surfactant topics is emphasized by the two common articles in Tables 23 and 24.

According to Tables 23 - 24 capillary pressure, wettability alteration, imbibition tests, carbonate minerals, polymeric surfactant, chemical enhanced oil recovery, brine salinity, fluid-fluid interaction, low salinity waterflooding, rheological analysis, nanoparticles as foam stabilizer, and foam flooding are terms revealing the Wettability Alteration and Surfactant topic s in the Oil Recovery cluster.

The Wettability Alteration topic is also relevant to the Crude Oil cluster, see Table 25.

Table 25. The most cited articles in the cluster level 1 Crude Oil and the cluster level 2 Wettability Alteration

TI	TC
The effect of brine salinity and oil components on dynamic IFT behavior of oil-brine during low salinity water flooding: Diffusion coefficient, EDL establishment time, and IFT reduction rate	11
Investigation into fluid-fluid interaction phenomena during low salinity waterflooding using a reservoir-on-a-chip microfluidic model	9
Microfluidic investigation of enhanced oil recovery: The effect of aqueous floods and network wettability	8
Experimental investigation on the dominating fluid-fluid and rock-fluid interactions during low salinity water flooding in water-wet and oil-wet calcites	7

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

Brine salinity, low salinity water flooding, network wettability, fluid-fluid and rock-fluid interactions, water-wet and oil-wet calcites — these terms correspond with the highly cited works in Tables 25 and fit well with the previous data in Table 23 and 24.

Table 26. The most cited articles in the cluster level 1 Crude Oil and the cluster level 2 Low Salinity,Low-salinity Water

TI	TC
The effect of brine salinity and oil components on dynamic IFT behavior of oil-brine during low salinity water flooding: Diffusion coefficient, EDL establishment time, and IFT reduction rate	11
Pore scale visualization of fluid-fluid and rock-fluid interactions during low-salinity waterflooding in carbonate and sandstone representing micromodels	10
Investigation into fluid-fluid interaction phenomena during low salinity waterflooding using a reservoir-on-a-chip microfluidic model	9
Nanoparticles as foam stabilizer: Mechanism, control parameters and application in foam flooding for enhanced oil recovery	8

Low-salinity Water is a common topic throughout the publications in Tables 23-27.

Table 27. The most cited articles in the cluster level 1 Numerical Simulation and the cluster level 2 Gas Production

TI	TC
A transient two-phase flow model for production prediction of tight gas wells with fracturing fluid-induced formation damage	37
A new transient model for hydrate slurry flow in oil-dominated flowlines	6
Impact of well shut-in after hydraulic-fracture treatments on productivity and recovery of tight oil reservoirs	5
Rate transient analysis of infinite-acting linear flow by use of piecewise constant diffusivity coefficients	5

tight gas wells и ight oil reservoirs - места применений Numerical Simulation в виде transient two-phase flow model, transient model for hydrate slurry flow, Rate transient analysis of infinite-acting linear flow.

Table 28. The most cited articles in the cluster level 1 Machine Learning and the cluster level 2 Error

TI	TC
Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models	20
Accurate determination of permeability in carbonate reservoirs using Gaussian Process Regression	14
A data-driven shale gas production forecasting method based on the multi-objective random forest regression	13

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

TI	TC
Determination of bubble point pressure & oil formation volume factor of crude oils applying multiple hidden layers extreme learning machine algorithms	11

Why Error? In abstracts: prediction error; absolute error; mean squared error; model-error; Mean Magnitude Relative Error; - типичные вопросы Machine Learning.

Why does the term Error appear in the title of the Level 2 cluster? Prediction error, absolute error, mean square error, model-error, and mean magnitude relative error - are typical terms of the Machine Learning topic that appear in the abstracts.

Machine Learning applications of interest are: prediction of petroleum reservoir properties, accurate determination of permeability in carbonate reservoirs, shale gas production forecasting, determination of bubble point pressure, i.e. areas for which it is difficult to build a working physical and mathematical model, but we can collect data for learning machine algorithms.

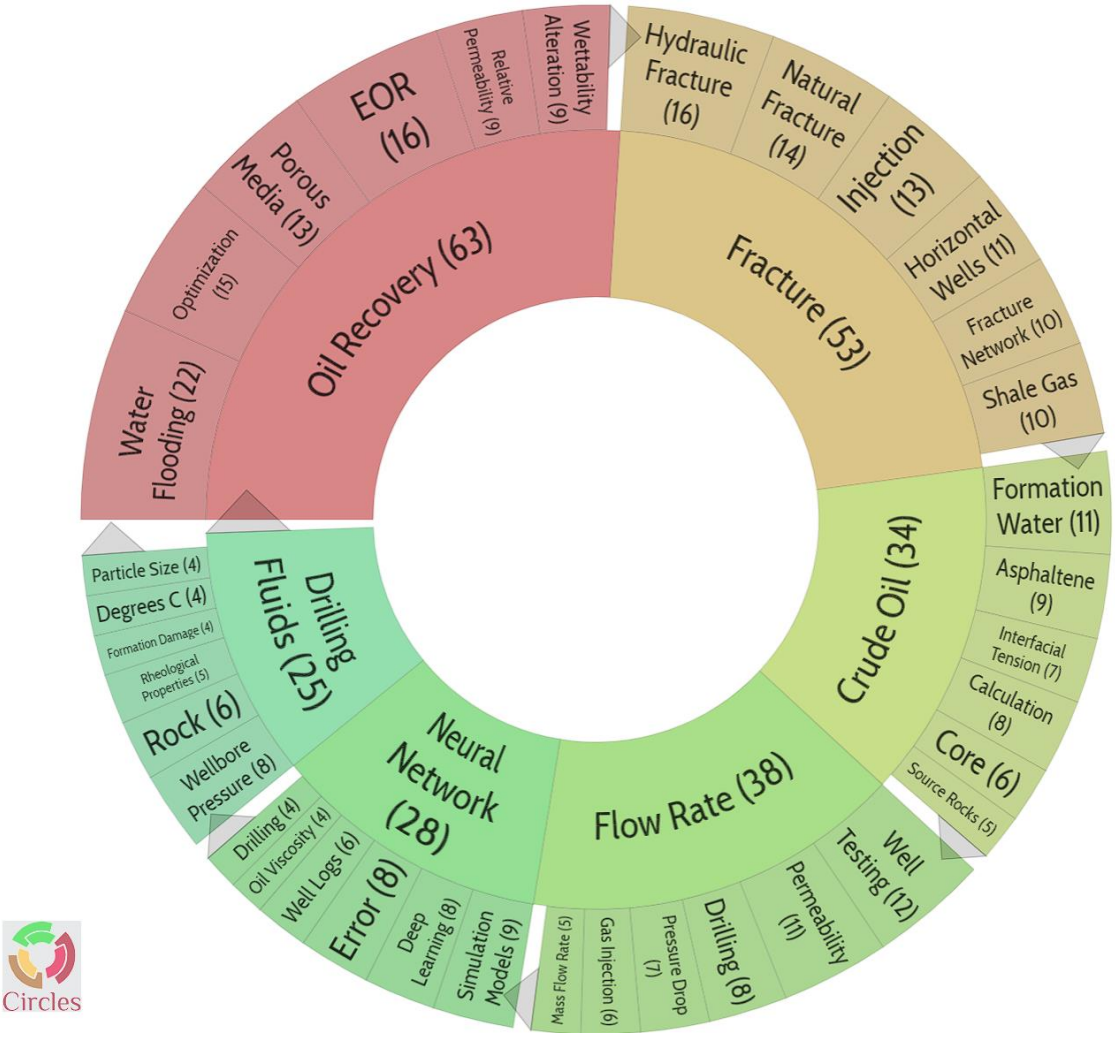

Fig. 8. Pie-chart for 350 least-cited articles in 2021.

The low-cited papers in Tables 29-30 have similarities with the highly cited papers in Tables 23-26 that study low-salinity waterflooding. The highly cited papers focus on wetting changes,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

surfactants, and more comprehensive treatment of low salinity, capillary pressure, and wetting changes, while the papers in Tables 29-30 are more general.

Table 29. The least-cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Water Flooding

TI	TC
Fluid dynamics analysis and performance of polymer flow regulators for polymer flooding in multilayered reservoirs	0
Modelling the impact of Alkaline-surfactant and Alkaline-surfactant-polymer flooding processes on scale precipitation and management	0
Forecasting the impact of formation damage on relative permeability during low-salinity waterflooding	0
Explicit continuum scale modeling of low-salinity mechanisms	0

Table 30. The least-cited articles in the cluster level 1 Oil Recovery and the cluster level 2 Optimization

TI	TC
An innovative workflow for selecting appraisal area in low permeability greenfield development under uncertainties	0
Efficient tracking and estimation of solvent chamber development during warm solvent injection in heterogeneous reservoirs via machine learning	0
Fast optimization of packer locations in wells with flow control completions	0
Well and ICV management in a carbonate reservoir with high gas content	0

The topics of the articles in Tables 31 and 32 overlap significantly with similar topics considered for articles published in 2018 and 2019, which is likely the reason for the low citation rates.

Table 31. The least-cited articles in the cluster level 1 Fracture and the cluster level 2 Hydraulic Fracture

TI	TC
Representation of high resolution rock properties on a coarser grid for hydraulic fracture modeling	0
Multigrid pressure solver for 2D displacement problems in drilling, cementing, fracturing and EOR	0
Characterizing mineralization on low carbon steel exposed to aerated and degassed synthetic hydraulic fracture fluids	0
An efficient stimulated reservoir area (SRA) estimation method based on octree decomposition of microseismic events	0

Table 32. The least-cited articles in the cluster level 1 Fracture and the cluster level 2 Natural Fracture

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 23 May 2022 doi:10.20944/preprints202205.0306.v1

TI	TC
Wellbore storage removal in pressure transient analysis for gas wells	0
New insights into forced and free fall gravity drainage performance in a fractured physical model	0
Statistical analysis of geological factors controlling bed-bounded fracture density in heterolithic shale reservoirs: The example of the Woodford Shale Formation (Oklahoma, USA)	0
Effect of compaction and imbibition on benefits of drawdown management in shale oil production: Uncertainty in recovery driving mechanisms	0

The low citation rates for 2021 papers in Tables 29-32 are consistent with the hypothesis that as research tasks become engineering rather than science, their citation rates in scientific publications fall.

This assumption deserves a separate study for a broader range of topics and areas of research.

Conclusions

It was shown that a comparison of highly cited and low-cited publications within a single highly ranked journal with a large number of publications can be an effective way of identifying promising research areas.

The validity of this approach is confirmed by the fact that its results are consistent with the generally accepted statement that highly cited papers deal with more relevant and detailed problems for a given time period.

Clusters of articles on Oil Recovery and Fracture are well represented in the 2018-2020 publications in the Journal of Petroleum Science and Engineering. The more cited papers address Low Salinity, Heavy, Wettability Alteration, Porous Media, Interfacial Tension and Sufractant within the Oil Recovery cluster and Natural Fractures, Fracture Propagation, Fracture Networks, Horizontal Wells, Shale Gas, Tight Oil Reservoirs, Mechanical Properties, Low Permeability and even Propant for the Fracture cluster. Less cited papers are characterized by Water Injection, Water Flooding and Hydraulics Fracture. In some cases, it can be noticed that the topics related to less cited tasks belonged to highly cited ones in the past years.

In 2020, publication topics were more conservative, perhaps due to the COVOD-19 pandemic and global attention to the problem.

In 2021, interest in Numerical Simulation, Machine Learning, and Neural Network methods increased significantly, especially in the fields of Gas Production, Poros Media, and Well Logs research.

Remark: I considered this work mainly as a preliminary test of the possibility of using bibliometric metadata not only to identify promising research areas described by key terms, but also to identify the most interesting publications related to promising research topics. As I gather and analyze scholarly publications related to promising areas of research, I have determined that a pure term clustering is hard to use to gather the most relevant publications for the field of study. There are a huge number of articles published nowadays, and if you use,

for example, two key terms for search, the sample got by such a query can be very large. On the other hand, three or four well-chosen terms greatly reduce the sample size and increase the value of the selected articles for further reading. Key term clustering is often based on their co-occurrence (pairwise relationship), so it can be difficult to gather the necessary publications by compiling a query with more than two terms from the same cluster. This encouraged me to move to using document clustering techniques rather than keywords. This approach simplifies the collection of scientific articles relevant to the ongoing work.

Carrot2 and LingoG3 are the best document clustering tools at the moment, in my opinion. The availability of demo versions on the Internet simplifies their use for a preliminary assessment of the feasibility of moving from term clustering to document clustering. The LingoG3 algorithm works pretty well, but is proprietary, so in further studies I consider reasonable to move to both open algorithms and open text analysis frameworks, such as Quanteda R package or Julia TextAnalysis.jl, or, better yet, use separate utilities for each stage of analysis.

In my opinion, the results presented in this preprint are encouraging, but only preliminary, requiring more detailed investigation.

References

[1] Aktar MostA, Harun MB, Alam MdM. Science, Technology, and Innovation (STI) Policy for Sustainable Development. In: Leal Filho W, Azul AM, Brandli L, Lange Salvia A, Wall T, editors. Affordable and Clean Energy, Cham: Springer International Publishing; 2020, p. 1–12. https://doi.org/10.1007/978-3-319-71057-0_59-1.

[2] Chigarev BN. Heavy oil: some aspects of bibliometric analysis for identifying promising scientific research. PEMOGC 2018:32–9. https://doi.org/10.30713/1999-6942-2018-8-32-39.

[3] Diligensky N.V., Tsapenko M.V., Davydov A.N. Multicriterion methodology identifying promising lines of research perspective // Vestnik of Samara State Technical University. Technical Sciences Series. - 2011. - Vol. 19. - N. 4. - P. 26-33. https://www.elibrary.ru/item.asp?id=17781708

[4] Park I, Yoon B. Identifying Promising Research Frontiers of Pattern Recognition through Bibliometric Analysis. Sustainability 2018;10:4055. https://doi.org/10.3390/su10114055.

[5] Mund C, Neuhäusler P. Towards an early-stage identification of emerging topics in science—The usability of bibliometric characteristics. Journal of Informetrics 2015;9:1018–33. https://doi.org/10.1016/j.joi.2015.09.004.

[6] Beskaravajnaja EV, Kharybina TN. Searching for a Method to Optimize the Scientific Activities of the Institute Based on Bibliometric Analysis of Research Themes. Sci Tech Inf Proc 2021;48:87–96. https://doi.org/10.3103/S0147688221020039.

[7] Jee J, Shin H, Kim C, Lee S. Six different approaches to defining and identifying promising technology through patent analysis. Technology Analysis & Strategic Management 2021:1–13. https://doi.org/10.1080/09537325.2021.1934437.

[8] Klemiński R, Kazienko P. Identifying Promising Research Topics in Computer Science. In: Alhajj R, Hoppe HU, Hecking T, Bródka P, Kazienko P, editors. Network Intelligence Meets User

Centered Social Media Networks, Cham: Springer International Publishing; 2018, p. 231–41. https://doi.org/10.1007/978-3-319-90312-5_16.

[9] Park I, Lee K, Yoon B. Exploring Promising Research Frontiers Based on Knowledge Maps in the Solar Cell Technology Field. Sustainability 2015;7:13660–89. https://doi.org/10.3390/su71013660.

[10] Rudskaya I, Ozhgikhin I, Kryzhko D. Developing and Testing an Algorithm to Identify Future Innovative Research Areas in Digitalization Conditions (using a Medical-sector example). IJTech 2020;11:1213. https://doi.org/10.14716/ijtech.v11i6.4438.

[11] Vérez D, Borri E, Cabeza LF. Trends in Research on Energy Efficiency in Appliances and Correlations with Energy Policies. Energies 2022;15:3047. https://doi.org/10.3390/en15093047.