Journal homepage:

http://ijimct.journals.ekb.eg/

Online ISSN: 2682-2881 **Print ISSN:** 2682-2105

Original Research Article

A Comprehensive Literature Review on Data Science Researches throughout the Period from 2008 to 2018 in IEEE Database

Dr. Ahmed Ibrahim Shahin*.

Information systems Dept., Computer Sciences and Information Technology college, Majmaah University, Majmaah 11952, KSA,

Dr.Ahmed Wafaa Hamdy†

Information systems Dept., Computer Sciences and Information Technology college, Majmaah University, Majmaah 11952, KSA,

ABSTRACT

In the recent years, Data science discipline took an effective role in various fields due the tremendous increase in the amount of structured and unstructured data produced daily. In this research, researchers studied Data science research trends. Researchers conducted a systematic Literature review (SLR) depending on Kitchen ham and Charters' proposed methodology. We reviewed journals' articles in IEEE Database. 67Articles were identified and reviewed throughout the Period from 2008 to 2018. By using inclusion and exclusion criteria 50 papers were selected for reviewing to answer the four research questions by conducting a comprehensive analysis on Index terms (Keywords), Countries and Affiliations, Citations, Usage and Number of researches. An increase in researches interest in Data science discipline was concluded, especially in the researches that describe the applications of Data science that handle both business needs to fulfil people and organizations requirements and Industrial and agricultural applications needs.

ARTICLE INFO

Article history:

Received 2020-10-04

Accepted 2020-12-30

Keywords:

Data science,
Literature review,
IEEE Database,
SLR,
Systematic Litreature Review,
Data science sub disciplines

* Corresponding author:Email: ai.ahmed@mu.edu.sa

† Corresponding author:Email: a.abdelhamid@mu.edu.sa

ABSTRACT

لعب تخصص علم البيانات فى السنوات الأخيرة دورًا فعالاً فى مختلف المجالات والتخصصات بسبب الزيادة الهائلة فى كمية البيانات المنظمة وغير المنظمة التى يتم إنتاجها يوميًا. ركزت هذه الورقة البحثية على دراسة الاتجاهات البحثية للمقالات العلمية المنشورة فى موضوع علوم البيانات. تم عمل مراجعة علمية لمقالات المجلات المنشورة فى قاعدة البيانات العالمية IEEE باستخدام منهجية (SLR) Systematic Literature reviewوهى المنهجية التى اقترحها Kitchen ham and Charters.
وقد تم فحص قاعدة البيانات العالمية IEEE لرصد المقالات ذات الصلة بمصطلح علوم البيانات، ثم تم اختيار ومراجعة 67 مقالة طوال الفترة من 2008 إلى 2018. بعد استخدام معايير التضمين والاستبعاد ، وقع الاختيار على 50 ورقة بحثية للمراجعة للإجابة على 4 أسئلة بحثية من خلال إجراء تحليل شامل للمصطلحات الكشفية (الكلمات الرئيسية) ، والبلدان، والجهات البحثية ، والاستشهادات المرجعية ، ومعدلات الاستخدام، وعدد الابحاث.
وكانت ابرز نتائج الورقة البحثية أن هناك زيادة فى اهتمام الباحثين بتخصص علم البيانات ، لا سيما فى الأبحاث التى تصف تطبيقات علوم البيانات التى تتعامل مع كل من احتياجات قطاع الأعمال لتلبية متطلبات الأشخاص والمنظمات واحتياجات التطبيقات الصناعية والزراعية.

علم البيانات؛
قاعدة بيانات IEEE ؛
التخصصات الفرعية لعلوم البيانات.

Introduction

Due to the increasing volume of recorded data by human race throughout many decades, Data Science discipline has emerged as a new discipline to handle tremendous amounts of data collected about everything at any time, and at anywhere. (Hilbert & Lopez, 2011)

Data science uses processes, methodologies and algorithms to extract knowledge from structured and unstructured data (Dhar, 2013). The main role of Data science is to interpret phenomena and introduce an analysis for it, using data analysis techniques, machine learning and other related methods. (Hayashi et al., 2015).

The main goal of Data science is to manipulate different data sources to figure out answers for the inquires listed below: (Aalst, 2014)

- What occurred?
- The reason why did it occur?
- The result of what happened?
- What is the recommended scenarios?

Due to the huge amounts of data emerging tremendously in various knowledge areas covering science fields from social sciences to healthcare, which led to computational resources overhead to process these gigantic amounts of data. Data science can be used to handle these situations as it focuses on problems emerging from huge amounts of data and its processing. (Simperl, 2017)

Data Science sub disciplines in literature review

Data is tremendously increasing in complexity in addition to being created, accessed and analyzed in real time. Data science extends the opportunity of implementing new approaches to ideas that contribute in providing more flexible methodologies. (Simperl, 2017)

Data Science combines techniques from statistics, databases, ethics, machine learning, data mining, high performance computing and visualization. This build Data science discipline relation to a large number of related disciplines including Process science. (Aalst, 2014)

Data science is based on statistics, assisted by Machine learning and Artificial intelligence techniques. (Mattmann, 2013).

Methodology and Scope

A systematic Literature review (SLR) was conducted with respect to the methodology introduced by Kitchen ham and Charters (Champiri et al., 2015). All (SLR) phases (planning, conducting, and reporting) were applied as shown in the next sections.

3.1. Phase 1: planning the review

In this phase, the main aim of the review was defined, and the following activities were performed.

3.1.1. Identification the need for review

In Phase 1, researchers identified that there was no clear evidence of availability of SLR in the discipline of Data science research trends in IEEE database. Due to the huge increase in number of studies describing the application of Data science in various humanity and applied fields, it is evident that applying this method in a different context could be considered a critical issue. So, determining the main research areas and trends of Data science would be an added value in directing future researches. Hence, researchers determined the necessity to perform a SLR according to the results of the previous studies.

3.1.2 Identifying Research Questions:

The main objective of this research is to explore and analyze research articles published in the period of 2008 to 2018 in IEEE database related to Data science. In order to achieve a more detailed and comprehensive view of this topic, the main objective was divided into next research questions.

RQ1- What are data Science Sub disciplines?
RQ2- What are the applications of Data science in various fields?
RQ3- What are the research trends in Data science in the period from 2008 to 2018?
RQ4- What are the leading countries in Data Science researches?

3.1. 3 Database Selection

According to research questions, IEEE database was selected to apply the research criteria using the IEEE Xplore digital library. (SDL, 2019)

3.2. Phase 2: Conducting the review

The research community and sample were identified through the following five steps:

First step, an advanced search on "Data Science" in IEEE Database was hold using IEEE Xplore according to the following criteria to retrieve full information:

- Search: Search in Meta data only.
- Searched item location in researches: Search with "Data Science" in document titles.
- Filters Applied: Journals & Magazines, Conferences, Early Access Articles.
- Period: Available period on IEEE database from the beginning of 1967 till the 15Th of March.

Second step, narrowing the results obtained from the first step was applied on research criteria to identify the research community as follows:

- Filters Applied: Journals & Magazines
- Period: Available period on IEEE database from the beginning of 1967 till 15Th of March 2018.

Third step, narrowing the results obtained from the second step was applied on research criteria to identify the research sample:

- Period: The period from the first of January 2008 till the end of December 2018.

Fourth step, a two levels analysis was conducted to measure the sample representing percent of the community.

Level 1: A comparison based on publishing media in IEEE Database showed 70.53% percent representing as shown in table 1 and fig. 1.

Table 1. Publishing media in IEEE Database in the periods 1967-2019 and 2008- 2019

Publishing media in IEEE Database	1967-2019	2008-2018	Percentage
Conferences	485	387	79.79%
Journals & Magazines	95	67	70.53%
Early Access Articles	5	2	40%
Books	12	10	83.33%

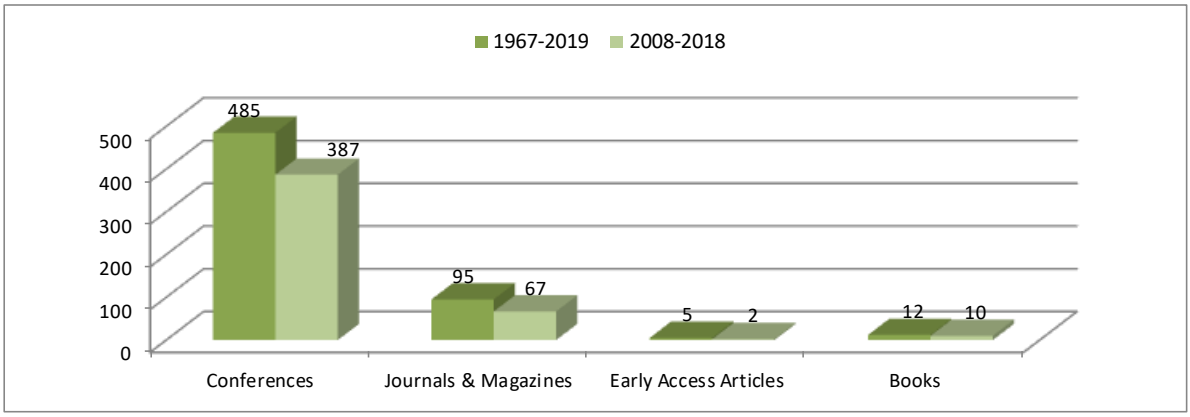

Fig. 1. Publishing media in IEEE Database in the periods 1967-2019 and 2008- 2019

Level 2: A comparison based on variables of Authors, Affiliations, Publication title, Publishers and Index terms (Keywords) showed true representing of the sample to the community as detailed in table 2 and fig. 2.

Table 2. Number of researches produced in the field of Data Science in the periods 1967-2019 and 2008- 2019

S	Item	1967-2019	2008-2018	Percentage
1	Authors	43	42	97.67%
2	Affiliation	35	31	88.57%
3	Publication Title	84	60	71.43%
4	Publisher	95	67	70.53%
5	Index terms (keywords)	133	94	70.7%

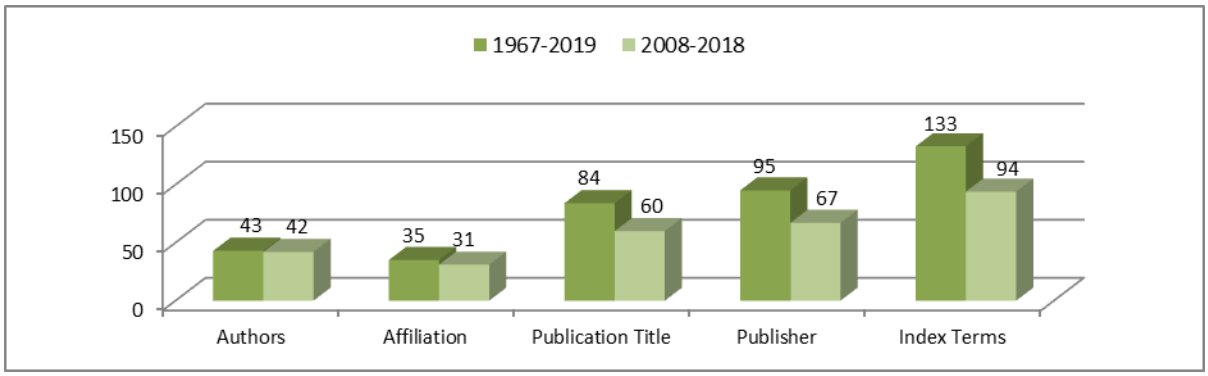

Fig 2. Number of researches produced in the field of Data Science in the periods 1967-2019 and 2008- 2019

Fifth step, a review was conducted on the resulted 67 researches contained in the research sample. Researchers reviewed the abstract, introduction, conclusion and keywords of the resulted papers. We selected those that:

- Articles introducing Data science as a discipline.
- Articles introducing Data science sub disciplines.
- Articles introducing Data science related disciplines.
- Articles introducing Data science application fields.

Furthermore, researchers determined some exclusion criteria as follows:

- Articles that contained the words “Data” & “Science” separated in researches’ document titles, and didn’t show relation to research scope (11 Articles).
- Articles introducing calls for papers (4 articles).
- Articles introducing Conference reports (2 articles).
-

As a result of the above review, the approved research sample consisted of 50 relevant articles (Appendix A) located in 27 journals (Appendix B).

Sixth step, in order to collect pertinent data from the selected articles, a data extraction method was applied to answer the previously listed study questions.

In order to find answers to research questions, researchers applied different strategies on the extracted data. Totally a narrative synthesis method was used, in addition to visualization methods like tables and charts based on research questions.

Results and Discussion

In this section, SLR phase 3 "reporting the review" is thoroughly discussed. The research sample is analyzed based on research questions mentioned in the research methodology according to the following variables:

4.1 Keywords

IEEE Database identifies three types of keywords as follows: (IEEE Xplore, 2019)

- **Author Keywords**
- **IEEE Terms**
- **Index Terms:**
    - **INSPEC Controlled Terms**
    - **INSPEC Non-controlled Terms**

In the research sample, researchers proceeded as follows:

First, the keywords were gathered from the research articles sample.

Second, an elicitation procedure was performed to eliminate duplicated and meaningless keywords.

Third, the keywords were classified according to "UNESCO Thesaurus". (UNESCO Thesaurus, 2019)

Forth, the used Keywords (shown in table 3) were categorized to research and vocational areas in computing adopted by IEEE computer society. (IEEE computer society, 2019)

Fifth, a comprehensive analysis was performed on the keywords of the research sample based on the previous steps, and showed the following (full details described in Appendix C):

Table 3. Total count and percentage of keywords research areas

Research area	Count of keywords	Percentage
Software Applications	336	37.21%
Information Technology	327	36.21%
Software Engineering	61	6.76%
Operating Systems and Networks	55	6.09%
Computer Theory	43	4.76%
Artificial Intelligence	37	4.10%
Computer Architecture	35	3.88%
Computer Design and Engineering	9	1.00%
Total	903	100%

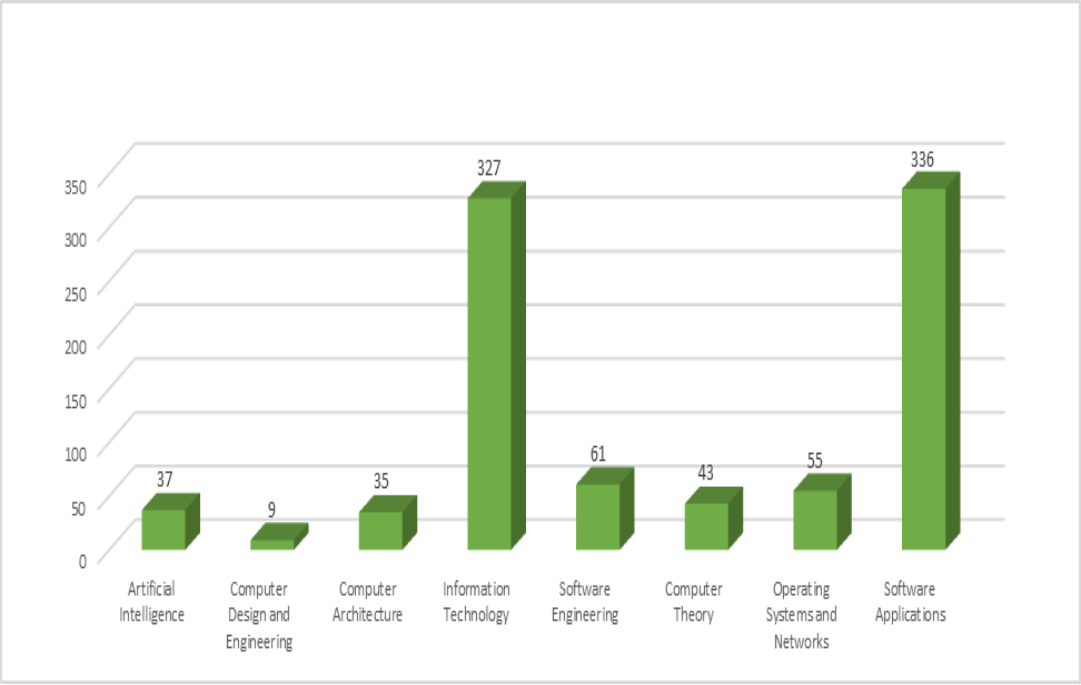

Fig 3. Total count of keywords research areas

The table 3 and fig. 3 shows the interest direction in Data science researches towards both Software applications (336 Keywords) which works to solve problems outside the computer field and Information technology (327 keywords) which serves business sector.

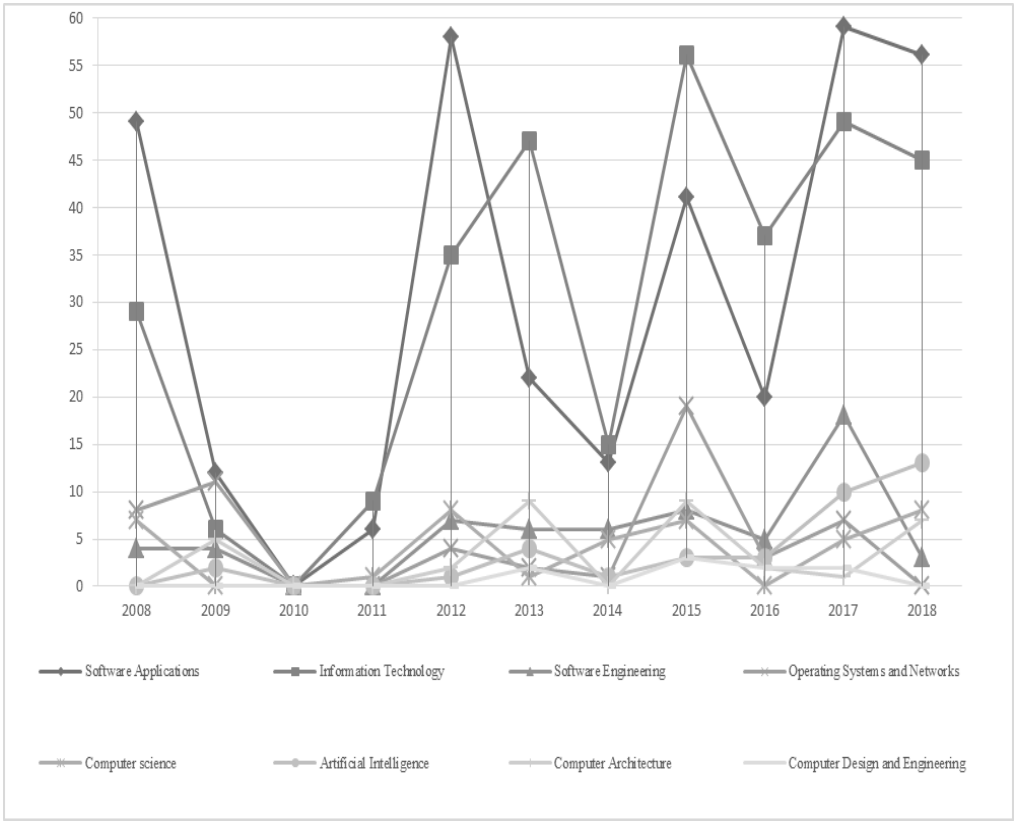

Fig 4. Count of keywords in research areas distributed among research sample duration (2008 - 2018)

Fig. 4 Show the pattern of change in research area as follows: (details in appendix C)

- Artificial Intelligence: the pattern shows a regular increase in the years 2009 and 2013, followed by a stability period from year 2015 to 2016, then a wide jump in the years 2017 and 2018. The pattern shows a pick every 4 years, this can be can be explained by the normal development duration for new trends in AI.
- Computer Design and Engineering: the pattern shows a weak appearance of Keywords starting from the year 2013, followed by a stability period from year 2015 to 2017, this can be explained by the indirect relation between research area scope and Data science.
- Computer Architecture: The pattern shows picks in the years 2009, 2013, 2015 and 2018. The average change pattern shows positive indication towards research interest, yet this interest doesn't show significant effect compared to other research areas.

- Information Technology: The pattern shows picks in the years 2013, 2015, 2017 and 2018. And Shows normal increase in years 2008, 2012 and 2016. The pattern identifies two change periods. The first period from 2008 to 2012 which show average start followed by high decrease till 2011 than a recovery in 2012. The second period from 2013 to 2018 show high records with narrow differences up and down except the year 2014 which showed low record. The pattern analysis shows an increasing interest in this research area.
- Software Engineering: the pattern shows a weak appearance of Keywords in general, the highest recorded keywords were in the year 2017. The remaining years shows near records. This can be explained by the indirect relation between research area scope and Data science.
- Computer Theory: the pattern shows a weak appearance of Keywords in general, It also shows a decrease in change period from 3 years to 2 years with conservation of the average keywords recorded. The interest in research area is normal due to its nature.
- Operating Systems and Networks: the pattern shows a weak appearance of Keywords in general, the highest recorded keywords were in the year 2015 and 2009. It shows pattern change increasing and decreasing on annual basis. This can be explained by the indirect relation between research area scope and Data science.
- Software Applications: The pattern shows picks in the years 2008, 2012, 2015, 2017 and 2018. And Shows normal increase in years 2013 and 2016. The pattern identifies four change periods. The first period from 2008 to 2012 which show high start followed by decrease till 2011 then a recovery in 2012.

  The second period from 2013 to 2015 showed a graduated decrease till 2014, then a recovery in 2015. The third period represented by year 2016 showed a significant decrease. The fourth period started with a recovery in 2017 followed by a non-remarkable decrease in the year 2018. The pattern analysis shows an increasing interest in this research area.
- From the previous analysis of the research sample keywords variable, researchers can conclude the presence of two main categories of research areas in relation with Data science in the sample as follows:
  - Research area with high interest: including Software Applications with 37.21% presence in the research sample and Information Technology with 36.21% presence in the research sample as shown in table 3. Fig. 5 emphasizes this result representing keywords density distributed among research areas.

- Research area with normal or low interest: including Software Engineering 6.76%, Operating Systems and Networks 6.09%, Computer Theory 4.76%, Artificial Intelligence 4.10%, Computer Architecture 3.88% and Finally Computer Science & Engineering 1%.

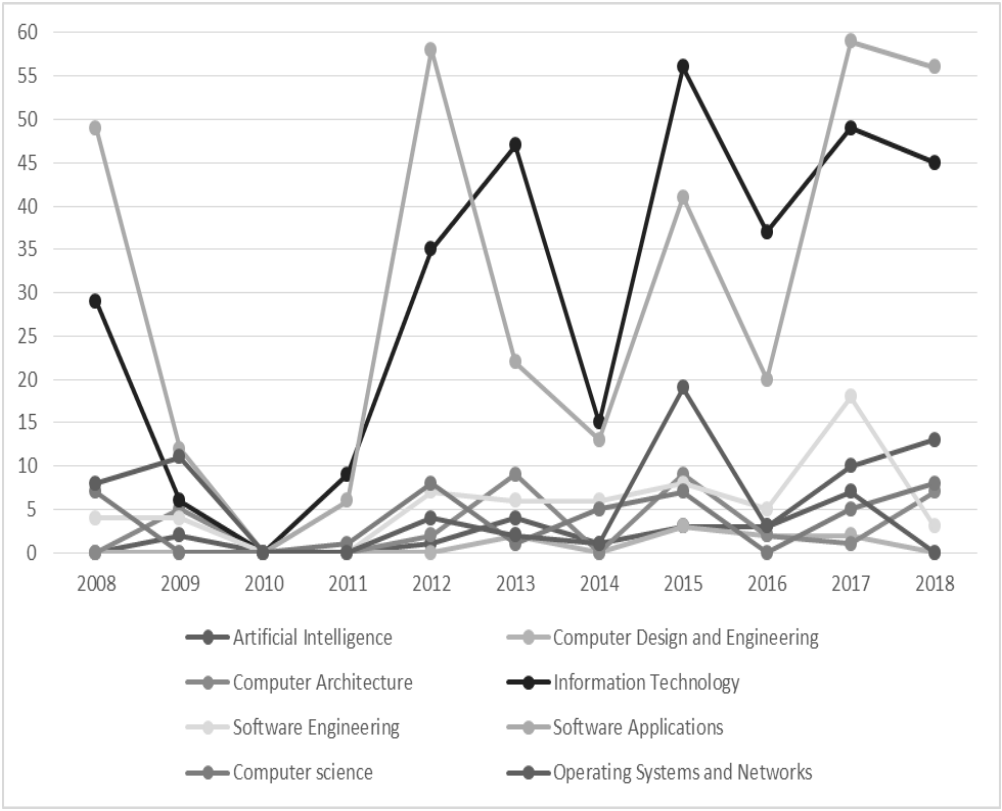

Fig 5. keywords density distributed among research areas

A through comparison was conducted between Software Applications and Information Technology to define the time line pattern for research interest in these two areas represented in fig. 6.

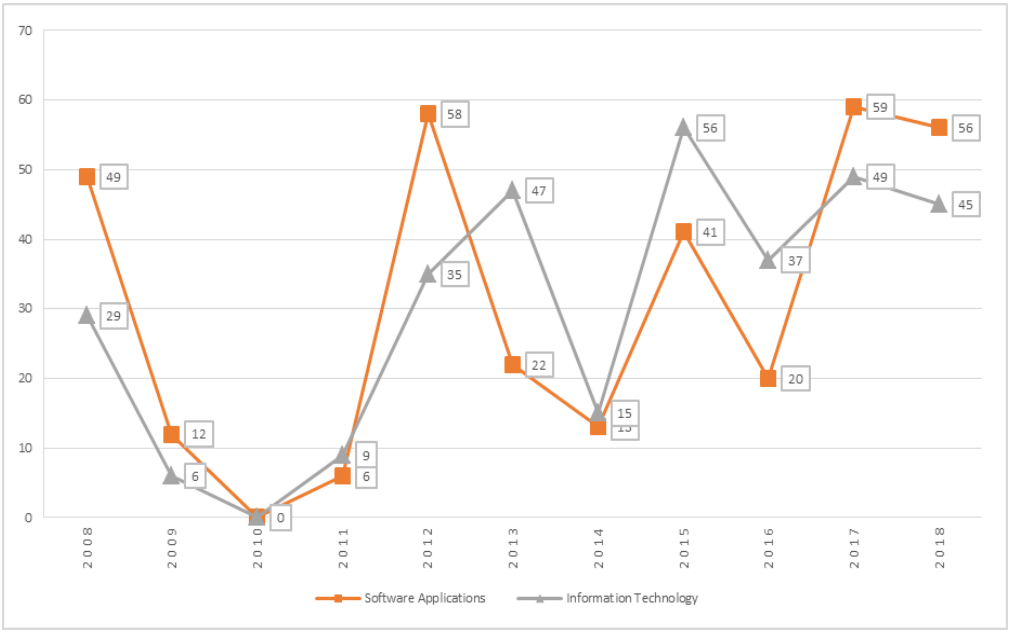

Fig 6. Top research areas distributed among research sample duration (2008 - 2018)

Fig. 6 shows great similarity in pattern development between the top two research areas, which indicates the direct relation between them within Data science research scope, as they both utilize Data Science; Information Technology to support business and organization and Software Applications to solve problems in various fields outside computer science.

Table 4. Keywords count in top 10 Research topics

Top 10 Research Topics	Keywords Counts	Research Area
Data science	139	Information Technology
Data processing	95	Information Technology
Medical	66	Software Applications
Computer Network	55	Operating Systems and Networks
Earth Science	43	Software Applications
Mathematics and Statistics	34	Computer Theory
Engineering	30	Software Applications
Information systems	28	Information Technology
Cloud Computing	28	Computer Architecture
General science	27	Software Applications

Table 4 shows the leading of "Data science" topic followed by "Data processing" topic that belong information technology area. In the third place came the "Medical" topic which belongs to Software Application area. In general 4 topics of the top 10 belong to Software Application research area, 3 topics belong to Information Technology, and the remaining 3 topics are distributed over other research areas. Researchers can conclude that research topics and research areas results have the same indication.

Table 5. Top ranked Keywords count among the research sample

Top ranked Keywords	Keywords Counts	Research Topics	Research Area
Data analysis	19	Data processing	Information Technology
Data Science	17	Data science	Information Technology
Big data	15	Big data	Information Technology
Scientific computing	12	Scientific computing	Software Applications
Data handling	11	Data processing	Information Technology
Data mining	11	Data Mining	Information Technology
Data models	9	Data science	Information Technology
Remote sensing	9	Computer Network	Operating Systems and Networks
Cloud Computing	8	Cloud Computing	Computer Architecture
Information retrieval	5	Information systems	Information Technology
Meteorology	5	Meteorology	Software Applications
Geophysics computing	5	Geoscience	Software Applications
Data visualization	5	Data processing	Information Technology

Table 5 shows the leading of "Data analysis", "Data Science" and "Big data" Keywords which belong to information technology area. In the fourth place came "Scientific computing" which belongs to Software Application area. In general, 8 Keywords belong to Information technology research area, 3 Keywords

belong to Software applications, and the remaining 2 Keywords are distributed on other research areas. The results confirm the leading of Information Technology and Software Applications research areas.

Fig. 7 illustrates the keywords development pattern throughout the research period. The years 2008, 2012, 2015 and 2017 showed an ascending count of keywords. As a result of Keywords Analysis throughout the research sample period, researchers noticed four years of pick keywords count which agreed with the results of keywords analysis in relation with research areas.

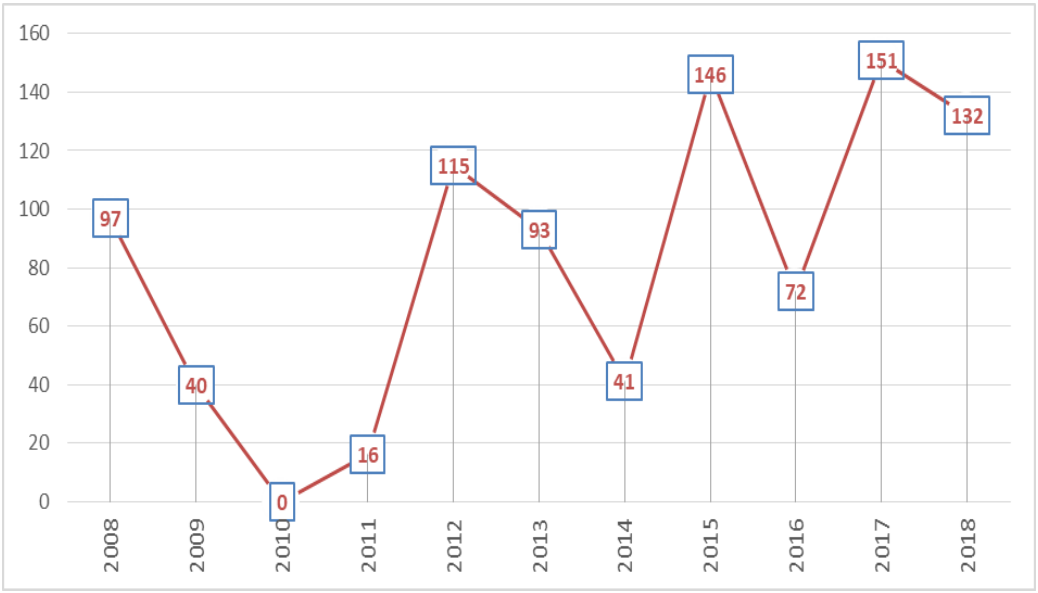

Fig 7. Keywords count throughout the research sample period (2008 - 2018)

By Analyzing keywords, researchers answered the following research Questions:

RQ1- What are data Science Sub disciplines?

RQ2- What are the applications of Data science in various fields?

RQ3- What are the research trends in Data science in the period from 2008 to 2018?

4.2 Countries and Affiliations

Institutional affiliation of authors listed in documents (university, government agency, corporation) (IEEE Xplore, 2019). According to IEEE Xplore statistics, which identified top 25 afflictions that produced Articles in Data science in our selected research sample. Researchers analyzed these affiliations to identify their home countries in order to obtain an indicator of the countries interest in data

science researches in our sample. The statistics of the sample results showed that USA is the first country in producing Data science researches with 65.3%, followed by China and Italy in the second place with 7.69% each, in the third place came Swiss, Japan, Korea, Brazil and UK with 3.85% each. Results are represented in fig. 8.

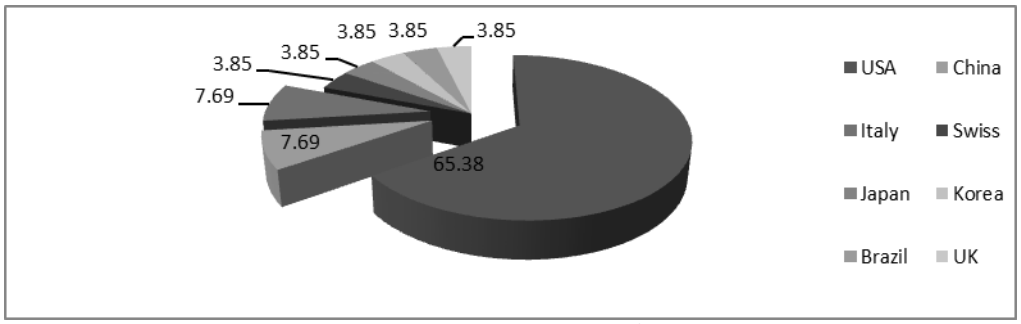

Fig 8. Countries that produced articles in Data Science in the period of 2008- 2019 within the research sample.

The Wide gap recorded in the research sample between USA and other countries especially China in articles number is due allowing publishing in English language only in IEEE database.

The results above answer the RQ4: What are the leading countries in Data Science researches?

4.3 Citations

The articles in the research sample were cited by 267 articles as described in table 6.

Table 6. Numbers and percentages of cited by researches for the research sample in the period from 2008 to 2018

S	Year	Cited by	Percent
1	2008	68	25.47%
2	2009	25	9.36%
3	2010	0	0%
4	2011	14	5.24 %
5	2012	15	5.62%
6	2013	46	17.23%
7	2014	10	3.75 %
8	2015	24	8.99%

9	2016	36	13.48%
10	2017	25	9.36%
11	2018	4	1.5%
Total		267	100%

- The highest number of cited in papers were registered in the year 2008 with a total of 68 and 25.47% of the total number of cited in papers during the research time interval.

- The fourth quarter of 2008 registered the highest number of cited in papers (59) during the research time interval. The research "TiSeG: A Flexible Software Tool for Time-Series Generation of MODIS Data Utilizing the Quality Assessment Science Data Set" gained the Whole cited by researches. This shows the trend direction of researches interest towards the applications of Data science. This answers RQ2- What are the applications of Data science in various fields?

Fig. 9 describes the pattern of increase and decrease of cited by researches, the pattern showed logical increase of number of cited by researches in the year 2008. The pattern also shows an average of 2 years change period both in increasing or decreasing direction of cited by researches.

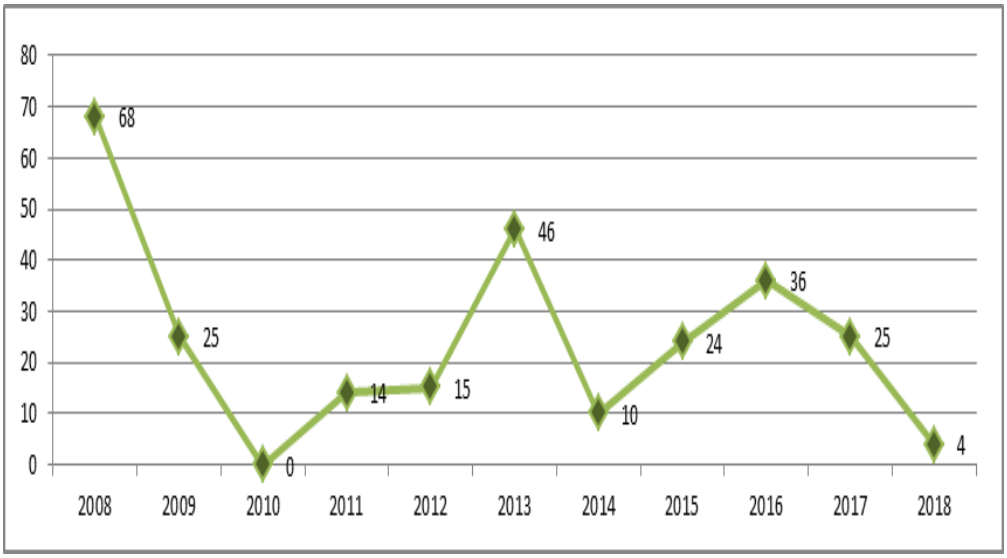

Fig 9. Cited by researches pattern in research sample in the period from 2008 to 2018

4.4 Usage

Usage gives an indicator of users' interest about Data science. The articles in the research sample showed a total number of 31785 Usage as described in table 7.

Table 7. Usage count and percentages for the research sample in the period from 2008 to 2018

S	Year	Usage	Percent
1	2008	1030	3.24%
2	2009	498	1.57%
3	2010	0	0%
4	2011	1016	3.2%
5	2012	2601	8.18%
6	2013	3369	10.16%
7	2014	2815	8.86%
8	2015	4673	14.70%
9	2016	8192	25.77%
10	2017	5617	17.67%
11	2018	1974	6.21%
Total		31785	100%

- The highest Usage was registered in the year 2016 with a total of 8192, and 25.77% of the total usage for papers during the research time interval.
- The third quarter of 2016 registered the highest usage (6704) during the research time interval. In this quarter, the research "Tell Me What You Eat, and I Will Tell You Where You Come From: A Data Science Approach for Global Recipe Data on the Web" gained the highest number of views (3586). This Show high interest in the applications of Data science that fulfil users' needs and interests. This answers RQ2- What are the applications of Data science in various fields?

Fig. 10 describes the pattern of increase and decrease in Usage during research sample period, the pattern showed logical increase in number of usage in the year 2016. The pattern also shows an average of 2 years change period in increasing or decreasing direction of usage. According to the pattern researchers could predict a new increase in users' interests reaches its pick in 2020.

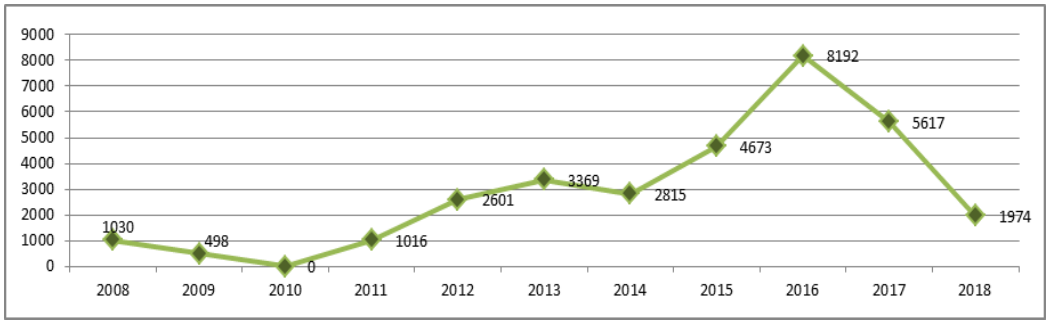

Fig 10. Researches usage pattern in research sample in the period from 2008 to 2018

4.5 Number of researches

The research sample contained 50 researches related to Data science discipline distributed among research period as shown in table 8.

Table 8. Researches count in the research sample in the period from 2008 to 2018

S	Year	Number of researches	Percent
1	2008	3	6%
2	2009	1	2%
3	2010	0	0%
4	2011	1	2%
5	2012	6	12%
6	2013	4	8%
7	2014	2	4%
8	2015	9	18%
9	2016	7	14%
10	2017	9	18%
11	2018	8	16%
	Total	50	100%

- The highest number of researches produced throughout the research period was registered in the years 2015 and 2017 with a count of 9 researches and 18% each.

Fig. 11 describes the pattern of increase and decrease of researches production during research sample period, the pattern shows increase in number of produced researches in the years 2015 and 2017. The pattern also shows an average of 2 years change period until 2014 in increasing or decreasing direction, starting from 2015 the average changed to 1 year instead of 2 years. According to the pattern researchers could predict a new increase in research production in 2019 and a slight decrease in 2020.

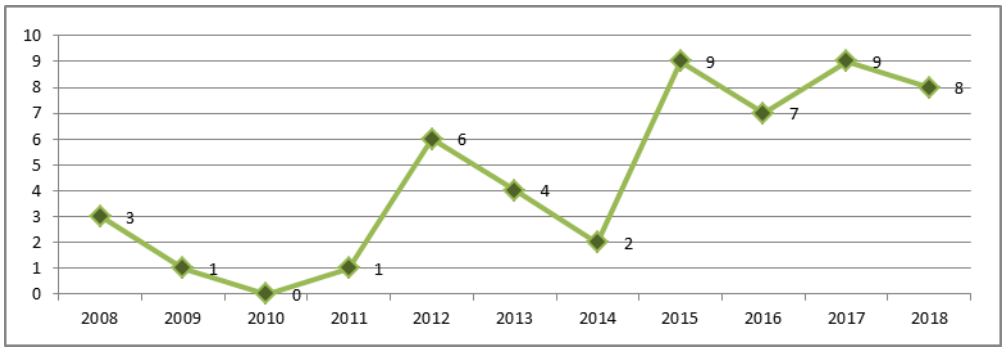

Fig 11. Researches production pattern in research sample in the period from 2008 to 2018

Conclusion

This systematic literature review (SLR) of Data Science Researches on IEEE Database has covered 67 papers published in 27 academic journals throughout the Period from 2008 to 2018. After using inclusion and exclusion criteria 50 paper were selected for the review to answer 4 research questions (RQ1 – RQ4) by conducting a comprehensive analysis on Index terms (Keywords), Countries and Affiliations, Citations, Usage and Number of researches.

As a result of "Keywords" analysis, researchers noticed the presence of two main categories of research areas in relation with Data science in the research sample that recorded high research interest, "Software Applications" with 37.2% and "Information Technology" with 36.2%. Research areas with normal or low interest included Software Engineering 6.8%, Operating Systems and Networks 6.1%, Computer Theory 4.8%, Artificial Intelligence 4.1%, Computer Architecture with 3.9% and Finally Computer Science & Engineering with 1%.

As a result of "Countries and Affiliations" analysis, USA is the first country in producing Data science researches with 65.3%, followed by China and Italy in the second place with 7.69% each, In the third place came Swiss, Japan, Korea, Brazil and UK with 3.85% each. The Wide gape recorded in the research sample between USA and other countries especially China in articles number is due allowing publishing in English language only in IEEE database.

As a result of "Citations" analysis, the pattern of increase and decrease of cited by researches showed logical increase in number of cited by researches in the year 2008. The pattern also showed an average of 2 years change period both in increasing or decreasing direction.

As a result of "Usage" analysis, the articles in the research sample Showed a total number of 31785 usage, the pattern showed logical increase of number usage in the year 2016. It also showed an average of 2 years change period in increasing or decreasing direction of usage. According to the pattern researchers could predict a new increase in users' interests reaches its pick in 2020. The researches sample analysis showed high interest in the applications of Data science that fulfil users' needs and interests.

As a result of "Number of researches" analysis, the highest number of researches produced throughout the research period in 2015 & 2017 with a count of 9 researches and 18% each. The pattern also showed an average of 2 years change period until 2014, starting from 2015 the average changed to 1 year. According to the pattern researchers could predict a new increase in research production in 2019 and a slight decrease in 2020.

Finally, researchers concluded an increase in researches interest in Data science discipline, especially in the researches that describe the applications of Data science that handle both business needs to fulfil people and organizations requirements and Industrial and agricultural applications needs.

Conflict of Interest

There is no conflict of Interest for this article.

References

Alan Fritzler, An Ethical Checklist for Data Science. Accessed Feb. 2019.- Avilable at: https://dssg.uchicago.edu/2015/09/18/an-ethical-checklist-for-data-science/

Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-2

C. Tankard, "Big Data Security", Network Security Newsletter, Elsevier, ISSN 1353-4858, July 2012

C.A. Mattmann, "Computing: A Vision for Data Science," Nature, vol. 493, no. 7433, 2013, pp. 473–475.

Chanchal Yadav, Shuliang Wang, Manoj Kumar. Algorithm and approaches to handle large Data- A Survey. IJCSN International Journal of Computer Science and Network, Vol 2, Issue 3, 2013. ISSN (Online) : 2277-5420

D.A. Keim ; F. Mansmann ; J. Schneidewind ; H. Ziegler, Challenges in Visual Data Analysis. Tenth International Conference on Information Visualisation (IV'06). 5-7 July 2006

Dhar, V. (2013). "Data science and prediction". Communications of the ACM. 56 (12): 64–73. doi:10.1145/2500499

C. Phethean, E. Simperl, T. Tiropanis, R. Tinati and W. Hall, "The Role of Data Science in Web Science," in IEEE Intelligent Systems, vol. 31, no. 3, pp. 102-107, May-June 2016.

Franck Cappello, Samir Djilali, Gilles Fedak, Thomas Herault, Frédéric Magniette, Vincent Néri, Oleg Lodygensky. Computing on large-scale distributed systems: XtremWeb architecture, programming models, security, tests and convergence with grid. Future Generation Computer Systems 21 (2005) 417–437

Han, Kamber, Pei, Jaiwei, Micheline, Jian (June 9, 2011). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann. ISBN 978-0-12-381479-

Hayashi, Chikio (1998-01-01). "What is Data Science? Fundamental Concepts and a Heuristic Example". In Hayashi, Chikio; Yajima, Keiji; Bock, Hans-Hermann; Ohsumi, Noboru; Tanaka, Yutaka; Baba, Yasumasa. Data Science, Classification, and Related Methods. Studies in Classification, Data Analysis, and Knowledge Organization. Springer Japan. pp. 40–51. doi:10.1007/978-4-431-65950-1_3. ISBN 9784431702085.

Iain Carmichae & J. S. Marron, Data science vs. statistics: two cultures?. Jpn J Stat Data Sci (2018) 1:117–138 https://doi.org/10.1007/s42081-018-0009-3

IEEE Xplore, Accessed Feb. 2019. Available at https://ieeexplore-ieee-org.sdl.idm.oclc.org/Xplorehelp/#/ieee-xplore-training/working-with-documents#working-with-abstracts

IEEE Computer Society. Careers in Computer Science and Computer Engineering. Accessed: 1st Feb. 2019, Available at: https://www.computer.org/communities/student-activities/career

Laura Cattaneo, Luca Fumagalli, Marco Macchi, Elisa Negri. Clarifying Data Analytics Concepts for Industrial Engineering. FAC-PapersOnLine, 2018 – Elsevier. Volume 51, Issue 11, 2018, Pages 820-825

M. Hilbert and P. Lopez, "The world's technological capacity to store, communicate, and compute information," Science, vol. 332, no. 6025, pp. 60–65, 2011. An example appendix

M. Smith, C. Szongott, B. Henne and G. Voigt, "Big Data Privacy Issues in Public Social Media", Digital Ecosystems Technologies (DEST), 6th IEEE International Conference on, Campione d'Italia, June 2012

Moustafa AA, Diallo TMO, Amoroso N, Zaki N, Hassan M, Alashwal H. Applying Big Data Methods to Understanding Human Behavior and Health. Front Comput Neurosci. 2018 Oct 16;12:84. doi: 10.3389/fncom.2018.00084.

S. Curry, E. Kirda, E. Schwartz, W. H. Stewart and A. Yoran, "Big Data Fuels Intelligence Driven Security", RSA Security Brief, January 2013 http://www. emc. com/collateral/industry-overview/big-data-fuels-intelligence- driven-security-io. Pdf

SDL. Access date: 1st Feb. 2019, Available at: https://sdl.edu.sa/SDLPortal/en/A-ZDataBases.aspx

UNESCO Thesaurus. Accessed: 1st Feb. 2019, Available at: http://vocabularies.unesco.org/thesaurus

W. v.d. Aalst, A. Adriansyah,"Process Mining Manifesto", Springer-Verlag Business Process Management Workshops ,vol. 99, 2011.

W. v.d. Aalst, "Data scientist: The engineer of the future," in Proc. I-ESA Conf., 2014, vol. 7, pp. 13–28. Authors including an appendix section should do so after References section. Multiple appendices should all have headings in the style used above. They will automatically be ordered A, B, C etc.

W. v.d. Aalst & Damiani, Ernesto. (2015). Processes Meet Big Data: Connecting Data. IEEE Transactions on Services Computing. 8. 1-1. 10.1109/TSC.2015.2493732.

Z.D. Champiri, S.R. Shahamiri, S.S.B. Salim, A systematic review of scholar context-aware recommender systems, Expert Syst. Appl. 42 (3) (2015) 1743–1758. https://www.sciencedirect.com/science/article/pii/S0957417414005569.

Appendices

Appendix A. Articles list in IEEE Database

Articles list
Processes Meet Big Data: Connecting Data Science with Process Science
Opportunities for Data Science in the Pharmaceutical Industry: The Use of Data to Find Efficiencies in Drug Development Can?t Come Too Soon
The Role of Data Science in Web Science
Data-Intensive Science
A Guide to Earth Science Data: Summary and Research Challenges
Big Data: Next-Generation Machines for Big Science
Noise-assisted data processing in measurement science: Part two
Data Science Data Governance
Data Treasure Hunters: Science Expanding to New Frontiers [In the Spotlight]
Big Data Analytics, Data Science and the CIS [President's Message]
Data Science: Nature and Pitfalls
Geographic Data Science
Science Data Management: Maximizing the Yield [Guest editors' introduction]
Guest Editorial Special Section on Fuzzy Systems in Data Science
Multimedia for Data Science
Analytical Intelligence in Processes: Data Science for Business
A Proposed Earth Science Collaboratory for Remote Sensing Data Analysis
Uncertainty-Aware Organ Classification for Surgical Data Science Applications in Laparoscopy
BlockPy: An Open Access Data-Science Environment for Introductory Programmers
Evolution of Information Management at the GSFC Earth Sciences (GES) Data and Information Services Center (DISC): 2006–2007
Data-Intensive Science in the US DOE: Case Studies and Future Challenges
Component- and Factor-Based Models for Data Fusion in the Behavioral Sciences
High-Accuracy Tidal Flat Digital Elevation Model Construction Using TanDEM-X Science Phase Data
Recent Activities in Earth Data Science [Technical Committees]
Applications of Computational Science: Data-Intensive Computing for Student Projects
Grand Challenge: Applying Regulatory Science and Big Data to Improve Medical Device Innovation
A Simple but Powerful Heuristic Method for Accelerating k -Means Clustering of Large-Scale Data in Life Science

Atmospheric Correction at AERONET Locations: A New Science and Validation Data Set
Dengue Epidemics Prediction: A Survey of the State-of-the-Art Based on Data Science Processes
Citizen Science with Hubble Space Telescope Data
Visualization of Citizen Science Volunteers' Behaviors with Data from Usage Logs
Data mining as a foundation for science-enabling autonomy
Pervasive Data Science
ADON: Application-Driven Overlay Network-as-a-Service for Data-Intensive Science
Big Data Challenges in Climate Science: Improving the next-generation cyberinfrastructure
Towards Data and Sensor Planning Service for Coupling Earth Science Models and Earth Observations
Cloud Paradigms and Practices for Computational and Data-Enabled Science and Engineering
SciDB for High-Performance Array-Structured Science Data at NERSC
The Next Grand Challenges: Integrating the Internet of Things and Data Science
Parallel Implementation of the Ensemble Empirical Mode Decomposition and Its Application for Earth Science Data Analysis
A Case for Data Commons: Toward Data Science as a Service
TiSeG: A Flexible Software Tool for Time-Series Generation of MODIS Data Utilizing the Quality Assessment Science Data Set
Data Prospecting–A Step Towards Data Intensive Science
Tell Me What You Eat, and I Will Tell You Where You Come From: A Data Science Approach for Global Recipe Data on the Web
Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data
Noise-assisted data processing in measurement science: Part one part 40 in a series of tutorials on instrumentation and measurement
The quest to save genomics: Unless researchers solve the looming data compression problem, biomedical science could stagnate
A New In-Car Navigation System Based on V2C2V and Data Science
The Future of Data-Intensive Science
Theory-Guided Data Science for Climate Change

Appendix B. Publishers in IEEE Database

Journal List
Aerospace and Electronic Systems Magazine, IEEE
Computer
Computing in Science & Engineering
IEEE Access
IEEE Cloud Computing
IEEE Computational Intelligence Magazine
IEEE Computer Graphics and Applications
IEEE Geoscience and Remote Sensing Magazine
IEEE Instrumentation & Measurement Magazine
IEEE Intelligent Systems
IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing
IEEE Latin America Transactions
IEEE MultiMedia
IEEE Pervasive Computing
IEEE Pulse
IEEE Security & Privacy
IEEE Signal Processing Magazine
IEEE Spectrum
IEEE Transactions on Biomedical Engineering
IEEE Transactions on Cloud Computing
IEEE Transactions on Fuzzy Systems
IEEE Transactions on Geoscience and Remote Sensing
IEEE Transactions on Knowledge and Data Engineering
IEEE/ACM Transactions on Computational Biology and Bioinformatics
IEEE TRANSACTIONS ON SERVICES COMPUTING
IT Professional
Proceedings of the IEEE

Appendix C. Count and percentage of keywords fields distributed among research areas

Research areas	Topics	Count	Total	%
Artificial Intelligence	Artificial intelligence	10	37	4.1%
	Machine learning	7		
	Computer vision	11		
	Image processing	9		
Computer Design and Engineering	Computer hardware	9	9	1.0%
Computer Architecture	Cloud Computing	28	35	3.9%
	Internet of Things	4		
	Computer architecture	3		
Information Technology	Management operations	14	327	36.2%
	Information Management	9		
	Information systems	28		
	Data processing	95		
	Big data	19		
	Data science	138		
	Data mining	24		
Software Engineering	Software application	17	61	6.8%
	Computer programming	25		
	Internet and Web application	19		
Computer Theory	Computer science	8	43	4.8%
	Mathematics and Statistics	35		
Operating Systems and Networks	Computer Network	55	55	6.1%
Software Applications	Food Industry	16	336	37.2%
	Engineering	30		
	Space Science	12		
	Astronomy science	10		
	Geoscience	13		
	Geography and oceanography	22		
	Earth Science	43		
	General science	27		
	Physics	3		
	Meteorology	7		
	Environmental degradation	10		
	Atmospheric Science	12		
	Natural sciences	11		
	Medical	67		
	Education	19		
	Social science	4		
	Behavioral science	6		
	Citizen science	4		
	Bioinformatics	8		
	Scientific computing	12		
	Total Keywords		903	100%

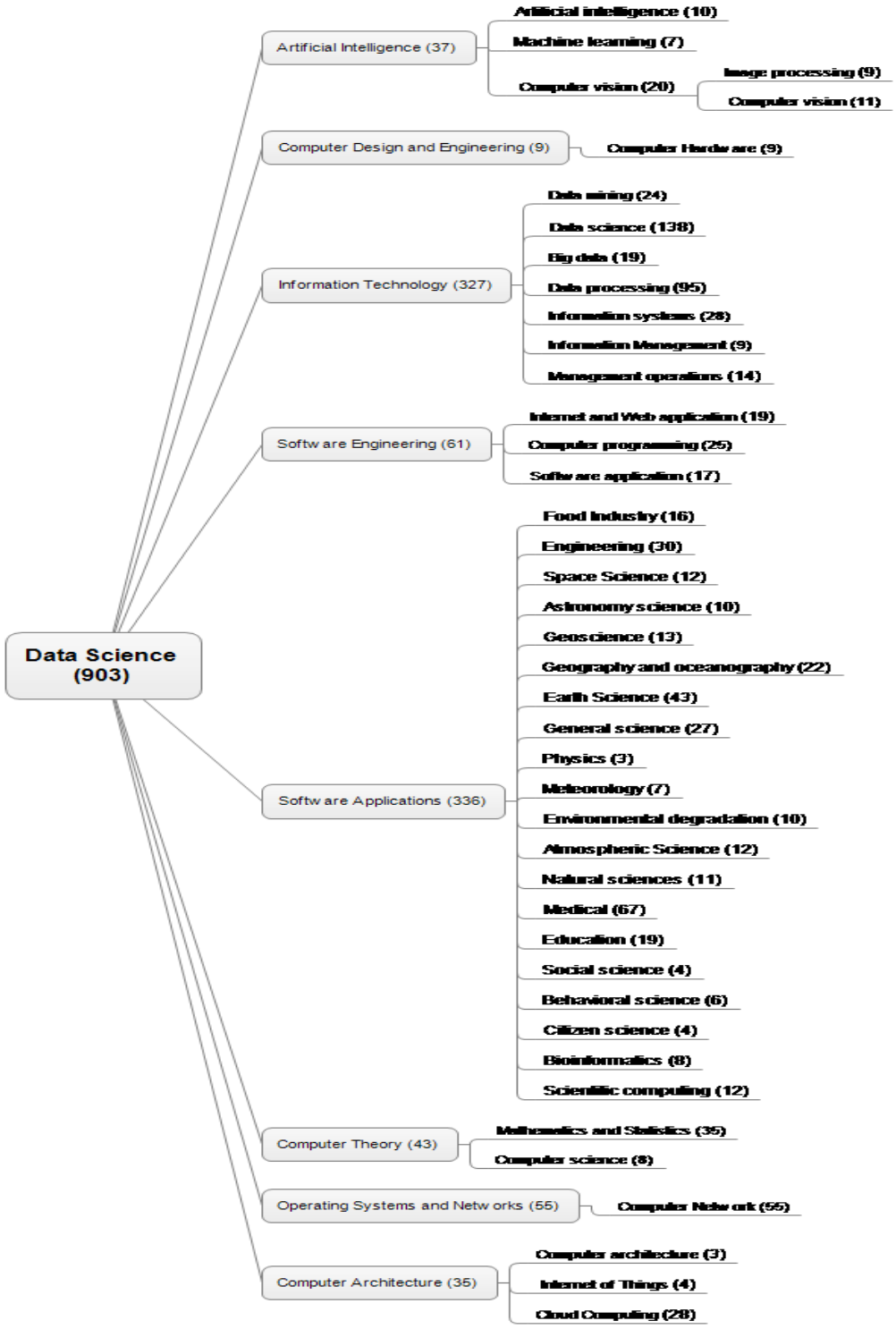

Fig C. Illustrates Count of keywords fields distributed among research areas

Appendix D. Count of keywords categorized by research areas distributed among research sample duration (2008 - 2018)

Years / Research areas	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Software Applications	49	12	0	6	58	22	13	41	20	59	56
Information Technology	29	6	0	9	35	47	15	56	37	49	45
Software Engineering	4	4	0	0	7	6	6	8	5	18	3
Operating Systems and Networks	8	11	0	0	4	2	1	19	3	7	0
Computer Theory	7	0	0	1	8	1	5	7	0	5	8
Artificial Intelligence	0	2	0	0	1	4	1	3	3	10	13
Computer Architecture	0	5	0	0	2	9	0	9	2	1	7
Computer Design and Engineering	0	0	0	0	0	2	0	3	2	2	0