Informatics for Health: Connected Citizen-Led Wellness and Population Health
R. Randell et al. (Eds.)

doi:10.3233/978-1-61499-753-5-73

Internet of Things in Health Trends Through Bibliometrics and Text Mining

Stathis Th. KONSTANTINIDIS[a,1], Antonis BILLIS[b], Heather WHARRAD[a], Panagiotis D. BAMIDIS[b]
[a] *School of Health Sciences, The University of Nottingham, United Kingdom*
[b] *Medical School, Aristotle University of Thessaloniki, Greece*

Abstract. Recently a new buzzword has slowly but surely emerged, namely the Internet of Things (IoT). The importance of IoT is identified worldwide both by organisations and governments and the scientific community with an incremental number of publications during the last few years. IoT in Health is one of the main pillars of this evolution, but limited research has been performed on future visions and trends. Thus, in this study we investigate the longitudinal trends of Internet of Things in Health through bibliometrics and use of text mining. Seven hundred seventy eight (778) articles were retrieved form The Web of Science database from 1998 to 2016. The publications are grouped into thirty (30) clusters based on abstract text analysis resulting into some eight (8) trends of IoT in Health. Research in this field is obviously obtaining a worldwide character with specific trends, which are worth delineating to be in favour of some areas.

Keywords. scientometrics, ubiquitous health, pervasive health, text data mining

1. Introduction

A new buzzword has come into the foreseen recently, namely Internet of Things. Kevin Ashton claim the first use of term Internet of Things (IoT) in 1999 [1] linking the idea of RFID in a supply chain. A lot of definitions are attempted with the Oxford dictionary defining IoT as "*the interconnection via the Internet of computing devices embedded in everyday objects, enabling them to send and receive data"*. The IEEE Internet Initiative definition as explained in [2] distinguishes between low and high complexity, thus it established separate definitions. Following the work of pervasive or ubiquitous computing, the Internet of Things in Healthcare is one of the key focus areas with examples including use in Active and Healthy Aging environments [3], "personalised preventative health coaches" [4], full body exergames to mobile devices [5] and others.

Bibliometric analysis provides a summary for research reported in scientific literature enabling researchers to generate quantitative information from existing data [6]. Text data mining can complement the bibliometric analysis and fulfil the need for faster content analysis and categorisation [7]. Text data mining or text mining involves information retrieval, text analysis, information extraction, clustering, categorization, visualization, database technology, machine learning and data mining [8].

[1] Corresponding author, DICE Research Group & HELM Team, School of Health Sciences, The University of Nottingham, Nottingham, NG7 2UH, UK; E-mail: Stathis.Konstantinidis@nottingham.ac.uk.

A recent study for Internet of Things using bibliometrics revealed [9] an increased number of publications in the last 5 years, while emerging and established research clusters were identified. Furthermore, there are a number of organisations which foresaw the rapid expansion of Internet of Things. The US National Intelligence Council include it in 2008 in a list of six "Disruptive Civil Technologies" with potential impact on US national power [10]. The UK government sees IoT as a transformative development which has enormous potential to change all of our lives [11] and foresees the potential areas for development of IoT in Healthcare to be prevention and early identification, research and tailored healthcare. The EU state that IoT represents the next step towards the digitisation of our society and economy and by 2020 the market size of IoT expected to exceed one trillion Euros [12].

Different organisations have different research priorities around Internet of Things in Health. Reviews on Internet of Things aiming to set the vision and identify the application and concepts have been made [13], [14], as well as an analysis via rigorous bibliometric and network analytics [9]. However, to the best of our knowledge there is no current research identifying the Trends of Internet of Things in Health through bibliometrics and text mining. This paper taking into consideration the field diversity in healthcare aims to answer the following questions: What are the IoT in Health bibliometrics? What are the themes in IoT in Health research? Are there any trends for IoT in Health research? What can be envisaged in the future in this domain?

2. Methods

Inspired by the work of Hung [7] we followed a similar methodology.

Data Collection: We chose the Web of Science as the source database since includes journals with highest impact in science and it is a bibliometric database which enables detailed bibliometric analysis. The following query was used to identify the relevant papers: *TS=("Internet of Things" OR IoT) AND (ehealth OR health* OR medic* OR nurs*) NOT (impairment oriented-training OR Integrated outpatient treatment OR Immunotech OR intravenous injection of endotoxin)* in order to include the term "Internet of Things" or the acronym IoT and one or more of words which relates IoT with health (e.g. ehealth, health, healthcare, medicine, medical, nurse, nursing, etc.) in the Abstract, Title and/or Keywords fields of a record. The acronym "iot" is used for different health related terms which excluded from the results. A total of 778 papers were retrieved. The search period was set from 1st of Jan 1998 to 30th Sep 2016. Despite the fact that not all the 2016 papers included, the latest papers might change the scene on the trends of internet of things in Health.

Data Analysis: The bibliometric data generated by the Web of Science downloaded locally and a set of bibliometric indicators were extracted, which describe the IoT in Health. BibExcel used to calculate co-occurrences of countries per paper and Pajek to visualise it. Text mining and clustering analysis performed using WordStat. The algorithm used is based on hierarchical clustering of key words (a text mining approach for automatic taxonomy generation and text categorisation), using as similarity metric the Adjusted Phi coefficient, a measure of association for two binary variables. In addition, the clustering method that was chosen was based on co-occurrence profiles (Second Order Clustering), considering that two keywords are close to each other, not necessarily because they co-occur but because they both occur in similar environments. Pre-processing involved stemming and exclusion of common words and phrases.

3. Results

Publication Data Collection: Figure 1(a) summarises the number of publication between 1998 and 2016. Publications of 2016 are included as they appear in Web of Science until September. The solid line represents the number of articles, while the dashed represents the moving average trend line. As the figure reveals the publications of IoT in Health are continuously growing. Document Type: The majority of the published literature in IoT in Health are proceedings papers while less than half are articles in journals. Twenty (20) are reviews while editorials, book chapters and abstracts are also present (Figure 1(b)).

Figure 1. IoT in Health : a) publication trends ; b) document types

The most prolific countries based on authors' affiliations are People's Republic of China (134) followed by USA (113), while UK (63), India (63), Spain (43) and South Korea (42) follow (Figure 2(a)). However, research is being conducted worldwide as figure 2(a) reveals with exceptions of Africa, a part of South America and middle Asia. The highest collaborations between countries is between People's Republic of China and USA, and as figure 2(b) depicts high cross-country collaboration exists. The most prolific source titles are LNCS (21), Sensors (20), Applied mechanics and materials (16) and Procedia Computer Science (15), while the 778 papers are published in 679 different sources.

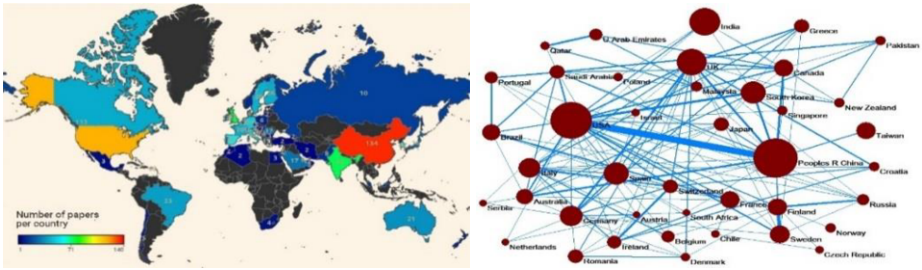

Figure 2. a) World research on IoT in Health based on authors' affiliation. b) Collaboration between countries based on authors' affiliations using BibExcel and Pajek.

Text Mining analysis through automatic clustering analysis created a dendrogram of 30 key term clusters. The 30 clusters interpret by two domain experts. Clusters with less than 5 articles and one cluster with noise (papers that were irrelevant to the topic) were dropped resulting to 745 articles. Since most of the terms exist in more than one article, frequency of cluster terms per article was used, to assign each article into a single cluster. As a result, a total number of eight (8) categories were formed by

combining several clusters together. As a further step, publishing trends were calculated for each category. Time trends are depicted in the Figure 3(b).

	Internet of Things in Health Trends	Articles
C1	Systems /Services design and implementation	258
C2	Communication/Connectivity Protocols & Algorithms	151
C3	Industrial potential of IoT	13
C4	Data Science analysis, storage and connectivity	92
C5	Quality Management and Privacy	23
C6	Efficiency and Cost of Application	172
C7	Smart Cities	15
C8	Ambient Assistive Living & Active Healthy Aging	21

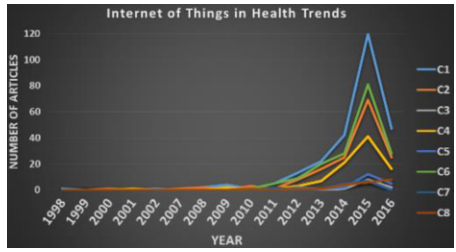

Figure 3. a) Trends identified and number of articles per trend for the selected period (Jan 1998 – Sep 2016). b) Time trends of articles identifying the growing interest of trends (number of articles per year).

The distribution of articles in C1, C2 and C6 trends follow the general worldwide research pattern, while C3 – "Industrial potential of IoT" trend apart from the 2 dominant Countries (Peoples Republic of China and USA (40% of the articles)) articles comes mainly from European countries (46% of the articles). C8 – "Ambient Assistive Living & Active Healthy Aging" is led by Greece, with the rest of the research spread worldwide. C4 – "Data Sciences analysis, storage and connectivity" which includes research on Big data, cloud computing and semantic web, is one of the few trends where Peoples Republic of China are not leading. Instead USA (20% of the articles), UK (15.7% of the articles) and India (10% of the articles) are the origin countries of this research topic. The C5- "Quality management and privacy" topic authoring country origin is led by European countries (43 % of articles), followed by Asia (30%) and North America (20%). Similar pattern applies for C7-"Smart Cities" with Europe to author 52.4% of the total articles followed by North America (23.8%) and Asia (20%).

4. Discussion and Conclusion

As the term 'Internet of Things' was claimed to be coined in 1999 we have selected the time range of the queries to be between 1998 and 2016. The growth of published research started in 2010, while a higher increase occurred after 2014 as Figure 1(a) and 3(a) depict. Since articles for 2016 are not included in full due to the time the search took place, the decreasing curve in the number of articles after 2015 is expected. The authors selected to include articles authored in 2016, since those formed differently the trends and allowed the latest focus of the published articles to be included.

The increased number of proceedings papers in comparison with journal articles (Figure 1(b)) might be an indicator for the rapid expansion of the field and the need for presenting early results and concepts, despite the fact that proceedings age faster they play a particularly important role in computer sciences (about 20% of the references)[15].

The institutional country of the authors reveals the countries where relevant IoT in Health research occurs (Figure 2 (a)) with the two dominant countries (Peoples Republic of China and USA) to have also the highest rate on collaboration between them through co-authorships (Figure 2(b)). As can been seen from Figure 2 cross-country collaborations are essential for this emerging field.

Some of the trends include more articles than others (Figure 3(a)). This could be linked with previous knowledge (e.g. C1, C2, C6) but also with research fields expansions such as big data and semantic web (C4), and the Active and Healthy Aging (C8). The time trends of articles identify a growing interest on the recognised topics. According to Figure 3(b), topics such as IoT industry potential (C3), quality management and privacy (C5) and smart cities (C7) are recently emerging trends since their numbers are increasing since 2014, while other more mature trends include ambient assisted living area (C8), systems design & implementation (C1), communication protocols (C2), data sciences (C4) and cost of implementation (C6), dating back in 2013 and 2011.

In this paper we formed eight (8) trends in Internet of Things in Health research through a bibliometric analysis. Research is worldwide with specific trends seen in certain countries. The findings of the study are limited and do not intended to be exclusive. This study might have missed articles in other scholar databases or articles that are in the process of publication. Different sets of keywords used in the search might generate different results that influence the trends of IoT in Health. This classification is by no means exhaustive and different classifications could be used. It is nevertheless already didactic from this piece of research that there exists already preferred domains of IoT applications which look not only promising for researchers but for facing societal challenges like active and healthy ageing as well.

References

[1] K. Ashton, "That 'Internet of Things' Thing," FRID J., 2009.
[2] R. Minerva, A. Biru, and D. Rotondi, "Towards a definition of the Internet of Things (IoT)," IEEE Internet Initiative. p.86, 2015.
[3] E. I. Konstantinidis, G. Bamparopoulos, A. Billis, and P. D. Bamidis, "Internet of things for an age-friendly healthcare.," Stud. Health Technol. Inform., vol. 210, pp. 587–91, 2015.
[4] D. V Dimitrov, "Medical Internet of Things and Big Data in Healthcare.," Healthc. Inform. Res., vol. 22, no. 3, pp. 156–63, Jul. 2016.
[5] E. I. Konstantinidis, G. Bamparopoulos, and P. D. Bamidis, "Transfering full body exergames from desktop applications to mobile devices: The role of the Internet of Things," in IMCL2015, 2015, pp. 254–258.
[6] M. Thelwall, "Bibliometrics to webometrics," J. Inf. Sci., vol. 34, no. 4, pp. 605–621, Jun. 2008.
[7] J. Hung, "Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics," Br. J. Educ. Technol., vol. 43, no. 1, pp. 5–16, Jan. 2012.
[8] A.-H. Tan, "Text Mining: The state of the art and the challenges," in Proceedings of the PAKDD 1999 Workshop on Knowledge Disocovery from Advanced Databases, 1999, vol. 8, pp. 65–70.
[9] D. Mishra, A. Gunasekaran, S. J. Childe, et al, "Vision, applications and future challenges of Internet of Things," Ind. Manag. Data Syst., vol. 116, no. 7, pp. 1331–1355, 2016.
[10] The National Intelligence Council and SRI Consulting Business Intelligence, "Disruptive Civil Technologies Six Technologies With Potential Impacts on US Interests Out to 2025," 2008.
[11] The Government Office for Science, "The IOT: making the most of the Second Digital Revolution," 2014.
[12] S. Aguzzi, D. Bradshaw, et al, "Definition of a Research and Innovation Policy Leveraging Cloud Computing and IoT Combination; A study prepared for the European Commission DG Communications Networks , Content & Technology - Contract number: 30-CE-0608166/00-79," 2014.
[13] S. Madakam, R. Ramaswamy, and S. Tripathi, "Internet of Things (IoT): A Literature Review," J. Comput. Commun., vol. 3, no. 3, pp. 164–173, 2015.
[14] G. Russo, B. Marsigalia, et al, "Exploring regulations and scope of the Internet of Things in contemporary companies: a first literature analysis" J. Innov. Entrep., vol. 4, no. 1, pp. 1–13, 2015.
[15] L. Cynthia, L. Vincent, and É. Archambault, "Conference proceedings as a source of scientific information: A bibliometric analysis" J. Am. Soc. Inf. Sci. Technol., vol. 59, no. 11, pp. 1776–1784, 2008.